Cu/e - H02712 - 60 - P014907

AMERICAN

60

JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

VOLUME LXVIII

1946

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

INDEX

PAGE

THE ASYMPTOTIC BEHAVIOR OF A SYSTEM OF LINEAR DIFFERENTIAL EQUATIONS.*

By Norman Levinson.

1. We shall prove the following result.

THEOREM I. *Let the linear system of differential equations with constant coefficients*

$$(1)\qquad y'_j(t) = \sum_{k=1}^{n} a_{jk} y_k(t) \qquad (j = 1, 2, \cdots, n),$$

have all of its solutions bounded as $t \to +\infty$. If the coefficients $f_{jk}(t)$ of the linear system

$$(2)\qquad x'_j(t) = \sum_{k=1}^{n} f_{jk}(t) x_k(t), \qquad (j = 1, 2, \cdots, n),$$

are continuous and satisfy

$$(3)\qquad \int^{\infty} | f_{jk}(t) - a_{jk} | \, dt < \infty, \qquad (j, k = 1, 2, \cdots, n),$$

then corresponding to any solution $x_j(t)$, $(j = 1, 2, \cdots, n)$, of (2) there is a solution $y_j(t)$, $(j = 1, \cdots, n)$, of the simple system (1) containing sinusoidal terms only, such that $x_j(t) - y_j(t) \to 0$ as $t \to +\infty$.

Theorem I generalizes a recent result of Wintner [4]. Wintner's result contains the added restriction that all the characteristic numbers of (1) must be purely imaginary, i.e. that the general solution of (1) consists of sums of pure sinusoids.

Here the general solution of (1) is of the form $y_j = \sum_{k=1}^{n} C_k \phi_{jk}$ where the C_k are arbitrary constants. Moreover there is an integer m, $0 \leqq m \leqq n$, such that

$$(4)\qquad \phi_{jk} = A_{jk} e^{i\lambda_k t}, \qquad (k = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n),$$

where the λ_k are real and the A_{jk} are constants. The terms of (4) are all sinusoids. The remaining terms in the solution of (1) are of the form

$$(5)\qquad \phi_{jk} = P_{jk}(t) e^{(-\beta_k + i\lambda_k)t}, \qquad (k = m+1, \cdots, n;\ j = 1, 2, \cdots, n),$$

* Received October 1, 1945.

where $\beta_k > 0$, and the $P_{jk}(t)$ are polynomials in t.

2. We shall assume in what follows that

$$\lambda_k \neq 0, \qquad (k = 1, 2, \cdots, m). \tag{6}$$

This is no real restriction. For if (6) is not satisfied the transformations $Y_j(t) = y_j(t)e^{-i\lambda t}$ and $X_j(t) = x_j(t)e^{-i\lambda_k t}$ with real $\lambda \neq \lambda_k$, $(k = 1, 2, \cdots, m)$, will transform (1) and (2) into new systems satisfying the hypothesis of Theorem I and such that for the new system (6) is satisfied. Proving Theorem I for the new systems will obviously yield the result for the original systems immediately.

We shall require

LEMMA 1. *The general solution of the system*

$$y'_j(t) = \sum_{k=1}^{n} a_{jk} y_k(t) + F_j(t), \qquad (j = 1, 2, \cdots, n),$$

can be written as

$$y_j = \sum_{k=1}^{m} \phi_{jk}(t) \int_{t_0}^{t} e^{-i\lambda_k \tau} \sum_{h=1}^{n} B_{kh} F_h(\tau)\, d\tau$$
$$+ \sum_{h=1}^{n} \int_{t_0}^{t} G_{jh}(t-\tau) F_h(\tau)\, d\tau + \sum_{k=1}^{n} C_k \phi_{jk}(t)$$

where the B_{kh} are constants, the C_k are arbitrary constants, and the $G_{jh}(t)$ satisfy

$$| G_{jh}(t) | \leqq M e^{-\beta t}, \qquad t \geqq 0, \tag{7}$$

where $\beta > 0$ and M are constants independent of j and h.

The proof of Lemma 1 will be given at the end of this paper.

Using the notation $f_{jk}(t) - a_{jk} = g_{jk}(t)$, (2) can be written as

$$x'_j(t) = \sum_{k=1}^{n} a_{jk} x_k + F_j \tag{8}$$

where

$$F_j = \sum_{k=1}^{n} g_{jk}(t) x_k(t). \tag{9}$$

Moreover from (3) we have

$$\int^{\infty} | g_{jk}(t) |\, dt < \infty. \tag{10}$$

To (8) we can apply Lemma 1 obtaining by also making use of (9),

$$x_j(t) = \sum_{k=1}^{m} \phi_{jk}(t) \sum_{h,s=1}^{n} B_{kh} \int_{t_0}^{t} e^{-t\lambda_k\tau} g_{hs}(\tau) x_s(\tau)\, d\tau + \sum_{h,s=1}^{n} \int_{t_0}^{t} G_{jh}(t-\tau) g_{hs}(\tau) x_s(\tau)\, d\tau + \sum_{k=1}^{n} C_k \phi_{jk}(t). \tag{11}$$

By proper choice of C_k any solution of (2) can be made to satisfy (11).

3. We shall now prove that the $x_j(t)$ are bounded as $t \to +\infty$. For the case where an n-th order linear differential equation with constants replaces (1) and an n-th order linear differential equation replaces (2) with a condition of the type (3) satisfied, Cesari [2] showed that the boundedness of the solutions of the simple system as $t \to +\infty$ implied that of the solutions of the perturbed system. A simplified proof of this has been given by Bellman [1]. That (3) is a "best possible" condition even if only boundedness is desired for the perturbed system follows from [3].

Using (4) and (7) we have from (11) that there exist an absolute constant M_1 and a constant $B > 0$ depending on the arbitrary constants C_k, such that

$$| x_j(t) | \leqq M_1 \sum_{h,s=1}^{n} \int_{t_0}^{t} | g_{hs}(\tau) x_s(\tau) |\, d\tau + B. \tag{12}$$

Setting

$$\sum_{h,s=1}^{n} | g_{hs}(\tau) | = H(\tau) \tag{13}$$

and

$$\max_{j=1,\ldots,n} | x_j(t) | = X(t)$$

we have from (12)

$$X(t) \leqq M_1 \int_{t_0}^{t} H(\tau) X(\tau)\, d\tau + B. \tag{14}$$

We also note that (10) implies

$$\int^{\infty} H(\tau)\, d\tau < \infty.$$

Clearly then we can choose t_0 so large that

$$M_1 \int_{t_0}^{\infty} H(\tau)\, d\tau < \tfrac{1}{2}. \tag{15}$$

(Of course the choice of t_0 may affect the magnitude of B but not of M_1.)

Now suppose that for some $t \geqq t_0$, $X(t) = 2B$. Then taking the smallest value of $t \geqq t_0$ for which this is true we have from (14) and (15)

$$2B < \tfrac{1}{2}(2B) + B = 2B$$

which is impossible. Thus $X(t) \leqq 2B$ and therefore

$$|x_j(t)| \leqq 2B, \qquad t \geqq t_0. \tag{16}$$

4. We now return to (11) and write

$$\int_{t_0}^{t} e^{-i\lambda_k\tau} g_{hs}(\tau) x_s(\tau)\, d\tau = \left(\int_{t_0}^{\infty} - \int_{t}^{\infty}\right) e^{-i\lambda_k\tau} g_{hs}(\tau) x_s(\tau)\, d\tau.$$

The first integral on the right is a constant. Modifying C_k, $(k = 1, 2, \cdots, m)$, to include the constants arising from these definite integrals we can write (11) as

$$x_j(t) = \sum_{k=1}^{m} C_k \phi_{jk}(t) + Q_j(t) + R_j(t) + S_j(t) \tag{17}$$

where

$$Q_j(t) = -\sum_{k=1}^{m} \phi_{jk}(t) \sum_{h,s=1}^{n} B_{kh} \int_{t}^{\infty} e^{-i\lambda_k t} g_{hs}(\tau) x_s(\tau)\, d\tau,$$

$$R_j(t) = \sum_{h,s=1}^{n} \int_{t_0}^{t} G_{jh}(t-\tau) g_{hs}(\tau) x_s(\tau)\, d\tau$$

$$S_j(t) = \sum_{k=m+1}^{n} C_k \phi_{jk}(t).$$

Using (4), (13) and (16) it follows that there exists a constant M_2 such that

$$|Q_j(t)| \leqq M_2 B \int_{t}^{\infty} H(\tau)\, d\tau.$$

Thus $Q_j(t)$ converges to zero as $t \to +\infty$.

Again, using (7) we have

$$|R_j(t)| \leqq 2BM \int_{t_0}^{t} e^{-\beta(t-\tau)} H(\tau)\, d\tau \leqq 2BM\, e^{-\frac{1}{2}\beta t} \int_{t_0}^{\frac{1}{2}t} H(\tau)\, d\tau + 2BM \int_{\frac{1}{2}t}^{t} H(\tau)\, d\tau.$$

Thus $R_j(t)$ converges to zero as $t \to +\infty$.

Finally by (5), $S_j(t)$ approaches zero as $t \to +\infty$. Thus (17) yields Theorem I.

5. We turn now to the proof of Lemma 1. This lemma results from representing the solution of

$$y'_j(t) = \sum_{k=1}^{n} a_{jk} y_k(t) + F_j(t), \qquad (j = 1, \cdots, n), \tag{18}$$

by what in operational calculus is called the superposition theorem.

We consider first the system

$$(19) \qquad \psi'_{jh}(t) = \sum_{k=1}^{n} a_{jk}\psi_{kh}(t) + \delta_{jh}, \qquad (j = 1, 2, \cdots, n),$$

where h is an integer, $1 \leqq h \leqq n$, and δ_{jh} is 1 if $j = h$ and is zero otherwise. A particular solution of (19) is given by $\psi_{jh} = d_{jh}$ where the d_{jh} are constants satisfying the equations

$$0 = \sum_{k=1}^{n} a_{jk}d_{kh} + \delta_{jh}, \qquad (j = 1, 2, \cdots, n).$$

The possibility of satisfying these equations results immediately from the fact that the determinant of $(a_{jk}) \neq 0$ which, in turn, is a consequence of (6). The general solution of (19) is given by

$$(20) \qquad \psi_{jh}(t) = \sum_{k=1}^{n} b_{kh}\phi_{jk}(t) + d_{jh}, \qquad (j = 1, 2, \cdots, n),$$

where the b_{kh} are constants. We choose the b_{kh} so that the $\psi_{jh}(t)$ satisfy the initial conditions

$$(21) \qquad \psi_{jh}(0) = 0, \qquad (j = 1, 2, \cdots, n).$$

We shall now show that a solution of the system (18) is given by

$$(22) \qquad y_j = \sum_{h=1}^{n} \int_{t_0}^{t} \psi'_{jh}(t - \tau)F_h(\tau)\,d\tau.$$

We have, on using (19) in (22),

$$y_j = \sum_{h,k=1}^{n} a_{jk} \int_{t_0}^{t} \psi_{kh}(t - \tau)F_h(\tau)\,d\tau + \int_{t_0}^{t} F_j(\tau)\,d\tau.$$

Differentiating and using (21) we have

$$y'_j = \sum_{h,k=1}^{n} a_{jk} \int_{t_0}^{t} \psi'_{kh}(t - \tau)F_h(\tau)\,d\tau + F_j(t).$$

Using (22) this becomes

$$y'_j = \sum_{k=1}^{n} a_{jk}y_k + F_j(t)$$

which is (18). Thus (22) is a particular solution of (18).

Using (20) and (4) we see that

$$\psi'_{jh} = \sum_{k=1}^{m} i\lambda_k b_k \, \phi_{jk}(t) + \sum_{k=m+1}^{n} b_{kh}\phi'_{jk}(t)$$

Recalling (5) and setting $\beta = \frac{1}{2} \min_{k>m} \beta_k$ we see that

$$\sum_{k=m+1}^{n} b_{kh}\phi'_{jk}(t) = G_{jh}(t)$$

where $G_{jh}(t)$ satisfies (7). Setting $i\lambda_k b_{kh} = B_{kh}$ we now have

$$\psi'_{jh} = \sum_{k=1}^{m} B_{kh}\phi_{jk}(t) + G_{jh}(t).$$

Using this in (22) we have, as the general solution of (18),

$$y_j = \sum_{k=1}^{m} \sum_{h=1}^{n} B_{kh} \int_{t_0}^{t} \phi_{jk}(t-\tau) F_h(\tau)\, d\tau$$
$$+ \sum_{h=1}^{n} \int_{t_0}^{t} G_{jh}(t-\tau) F_h(\tau)\, d\tau + \sum_{k=1}^{n} C_k\phi_{jk}(t).$$

Using (4) this yields the result of Lemma 1.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY.

REFERENCES.

[1] R. Bellman, "The stability of solutions of a differential equation," *Duke Mathematical Journal*, vol. 10 (1943), pp. 643-648.

[2] L. Cesari, "Sulla stabilità delle soluzioni delle equazioni differenziali lineari," *Ann. Scuola norm. super.* Pisa, vol. 8 (1939), pp. 131-148.

[3] N. Levinson, "The growth of solutions of a differential equation," *Duke Mathematical Journal*, vol. 8 (1941), pp. 1-10.

[4] A. Wintner, "Small perturbations," *American Journal of Mathematics*, vol. 67 (1945), pp. 417-430.

COMMENT ON THE PRECEDING PAPER.*

By HERMANN WEYL.

The following lines, though adding little to the substance of Professor Levinson's paper, may help to shed some light on his interesting result. This result is not limited to linear equations; it holds for such perturbations as are majorized by linear perturbations. Levinson passes from a given solution $\mathfrak{x}$ of the complete equation to a corresponding one $\mathfrak{z}$ of the approximate equation, $\mathfrak{x} \rightarrow \mathfrak{z}$. We add the inverse process $\mathfrak{z} \rightarrow \mathfrak{x}$: transition from the unperturbed to the perturbed phenomenon.

A system of differential equations for n functions $x_i(t)$ of the real variable t may be looked upon as a single differential equation for the vector (column) $\mathfrak{x}$ with the components x_i, and assuming the right member to consist of a linear part $A\mathfrak{x}$ with a constant coefficient matrix $A = \| a_{ik} \|$ and a perturbation $\mathfrak{b}$, we may write the equation in the form

$$(1) \qquad d\mathfrak{x}/dt = A\mathfrak{x} + \mathfrak{b}, \qquad \mathfrak{b}(t) = \mathfrak{f}(t, \mathfrak{x}(t)),$$

$\mathfrak{f}(t, \mathfrak{x})$ being a given vector function of t and a variable vector $\mathfrak{x}$. Form $U(t) = e^{At}$, the one-parameter group of linear transformations generated by the infinitesimal A:

$$dU/dt = AU = UA, \quad U(0) = \text{unit matrix } E; \quad U(t - \tau) = U(t)U^{-1}(\tau),$$

and define the absolute values $\| \mathfrak{x} \|$, $\| A \|$ of vectors and matrices by

$$\| \mathfrak{x} \|^2 = \Sigma \mid x_i \mid^2, \qquad \| A \|^2 = \Sigma \mid a_{ik} \mid^2.$$

The equation (1) may be written

$$(d/dt)(U^{-1}\mathfrak{x}) = U^{-1}\mathfrak{b}$$

("variation of constants"), whence follows

$$(2) \qquad \mathfrak{x}(t) = \mathfrak{z}(t) + U(t) \int_0^t U^{-1}(\tau)\mathfrak{b}(\tau)\,d\tau,$$

$$(3) \qquad \mathfrak{z}(t) = U(t)\mathfrak{a} \qquad (\mathfrak{a} = \text{const.})$$

* Received December 5, 1945.

being a solution of

$$\text{(4)} \qquad d\mathfrak{z}/dt = A\mathfrak{z}.$$

It can also be verified directly that if $\mathfrak{x}$ is a solution of (1), then the $\mathfrak{z}$ defined by (2) or

$$\text{(5)} \quad \mathfrak{z}(t) = \mathfrak{x}(t) - U(t) \int_0^t U^{-1}(\tau)\mathfrak{b}(\tau)\,d\tau = \mathfrak{x}(t) - \int_0^t U(t-\tau)\mathfrak{b}(\tau)\,d\tau$$

is a solution of $d\mathfrak{z}/dt = A\mathfrak{z}$. This is, in general form, the lemma from which Mr. Levinson's investigation starts.

Let us now assume that

$$\text{(I)} \qquad \| \mathfrak{f}(t, \mathfrak{x}) \| \leq \| \mathfrak{x} \| \cdot g(t).$$

If the perturbation is linear, $\mathfrak{f}(t, \mathfrak{x}) = G(t)\mathfrak{x}$, then (I) holds with $g(t) = \| G(t) \|$.

LEMMA. *Suppose*

$$\| \mathfrak{z}(t) \| \leq a, \qquad \| U(t) \| \leq c \quad \textit{for} \quad 0 \leq t \leq t_1$$

and set

$$c \int_0^t g(\tau)\,d\tau = h(t).$$

Then the equation (5) *implies*

$$\| \mathfrak{x}(t) \| \leq a \cdot e^{h(t)} \qquad (0 \leq t \leq t_1).$$

Proof. For $x(t) = \| \mathfrak{x}(t) \|$ one obtains, as a consequence of assumption (I), the integral inequality

$$x(t) \leq a + c \int_0^t x(\tau) g(\tau)\,d\tau = a + \int_0^t x(\tau) \cdot dh(\tau);$$

hence for $y(t) = x(t) - a \cdot e^{h(t)}$ the inequality

$$y(t) \leq \int_0^t y(\tau) \cdot dh(\tau).$$

Choose a constant b such that $y(t) \leq b$ for $0 \leq t \leq t_1$. The inequalities

$$y(t) \leq b \cdot (h(t))^n/n! \qquad (n = 0, 1, 2, \cdots;\ 0 \leq t \leq t_1)$$

then follow one after the other by induction. In the limit for $n \to \infty$ one gets $y(t) \leq 0$.

Let us now introduce Levinson's two hypotheses:

(II) convergence of $q = \int_0^\infty g(\tau)\,d\tau$;

(III) boundedness of $U(t)$ for $t \to \infty$, $\| U(t) \| \leq c$ for $t \geq 0$.

Then of course every solution $\mathfrak{z}$, (3), of (4) is bounded, $\| \mathfrak{z}(t) \| \leq a = c \| \mathfrak{a} \|$. Our lemma yields at once

$$\| \mathfrak{x}(t) \| \leq a^* = a \cdot e^{cq} \quad (\text{for } t \geq 0), \tag{6}$$

and hence the

THEOREM. *Under the hypotheses* (I), (II), (III) *every solution of* (1) *is bounded for* $t \geq 0$.

Instead of (6) Levinson uses the rougher estimate

$$\| \mathfrak{x}(t) \| \leq a/(1-\theta) \qquad (0 \leq t \leq t_1;\ \theta = h(t_1)),$$

which he establishes in a more indirect way. Its validity is limited to intervals $0t_1$ for which $\theta < 1$. Nonetheless, the theorem can be derived from it: one simply replaces the lower limit of integration 0 by a t_0 so chosen that

$$\theta = c \int_{t_0}^\infty g(\tau)\,d\tau < 1.$$

In a suitable coordinate system $U(t)$ breaks up into blocks (elementary divisors) of the form

$$\begin{Vmatrix} e^{\lambda t} \cdot t_0, & 0, & \cdots & 0 \\ e^{\lambda t} \cdot t_1, & e^{\lambda t} \cdot t_0, & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ e^{\lambda t} \cdot t_{m-1}, & e^{\lambda t} \cdot t_{m-2}, & \cdots & e^{\lambda t} \cdot t_0 \end{Vmatrix} \qquad [\ t_i = (1/i!)t^i\ ].$$

Hence hypothesis (III) implies that either $\Re\lambda < 0$ or $\Re\lambda = 0$; but in the latter case the block must consist of one row only ($m = 1$). After uniting all elementary divisors of the first and second kind into V_1, V_2 respectively, one gets a decomposition $U = \begin{Vmatrix} V_1 & 0 \\ 0 & V_2 \end{Vmatrix}$ in which $V_1(t) \to 0$ for $t \to \infty$ and $V_2(t)$ is bounded not only for $t \to +\infty$, but also for $t \to -\infty$. Using the corresponding decomposition of the unit matrix

$$E = \begin{Vmatrix} E_1 & 0 \\ 0 & 0 \end{Vmatrix} + \begin{Vmatrix} 0 & 0 \\ 0 & E_2 \end{Vmatrix} = I_1 + I_2.$$

into two complementary idempotents I_1, I_2, one can write

$$U = UI_1 + UI_2 = U_1 + U_2.$$

The idempotents I_1 and I_2 commute with A (and hence with U). In this form the description is independent of the coordinate system.

$$U_1(t) \to 0 \text{ for } t \to \infty, \qquad \| U_2(t) \| \leq c_2 \text{ for } t \leq 0.$$

Every vector $\mathfrak{x}$ may be split according to $I_1\mathfrak{x} + I_2\mathfrak{x} = \mathfrak{x}_1 + \mathfrak{x}_2$. For a solution $\mathfrak{z}$ of (4) the first part $\mathfrak{z}_1$ is damped, the second $\mathfrak{z}_2$ a pure sinusoidal oscillation.

Let us now decompose U in (5) into $U_1 + U_2$ and in the second part replace the lower limit of integration 0 by ∞:

$$\mathfrak{x}(t) - \int_0^t U_1(t-\tau)\mathfrak{v}(\tau)\,d\tau + \int_t^\infty U_2(t-\tau)\mathfrak{v}(\tau)\,d\tau = \mathfrak{z}(t). \tag{7}$$

Since $\mathfrak{x}(t)$ is bounded for $t \geq 0$, (6), and

$$\| \mathfrak{v}(\tau) \| = \| \mathfrak{f}(\tau, \mathfrak{x}(\tau)) \| \leq \| \mathfrak{x}(\tau) \| \, g(\tau) \leq a^* g(\tau), \text{ for } \tau \geq 0,$$
$$\| U_2(t-\tau)\mathfrak{v}(\tau) \| \leq c_2 a^* g(\tau) \text{ for } \tau \geq t \geq 0,$$

the integral extending to infinity converges by virtue of hypothesis (III). Observe that

$$U_2(t-\tau) = U(t-\tau)I_2 = U(t)U(-\tau)I_2 = U(t)U_2(-\tau).$$

The fact that $\mathfrak{z}(t)$, (7), satisfies (4) provided $\mathfrak{x}$ is a solution of (1) is not affected by the shift of the limit of integration in the U_2-part, as one sees either by direct verification or (as Levinson does) by the simple transformation

$$U(t) \int_0^t U_2(-\tau)\mathfrak{v}(\tau)\,d\tau = U(t)\mathfrak{a}_2 - U(t) \int_t^\infty U_2(-\tau)\mathfrak{v}(\tau)\,d\tau$$

where $\mathfrak{a}_2 = \int_0^\infty U_2(-\tau)\mathfrak{v}(\tau)\,d\tau$. Therefore formula (7) associates with every solution $\mathfrak{x}$ of (1) a unique solution $\mathfrak{z}$ of (4), $\mathfrak{x} \to \mathfrak{z}$. Multiplication by I_1, I_2 splits (7) into the two equations

$$\mathfrak{x}_1(t) - \int_0^t U_1(t-\tau)\mathfrak{v}(\tau)\,d\tau = \mathfrak{z}_1(t),$$

$$\mathfrak{x}_2(t) + \int_t^\infty U_2(t-\tau)\mathfrak{v}(\tau)\,d\tau = \mathfrak{z}_2(t),$$

and thus one arrives at

LEVINSON'S THEOREM. *For every solution $\mathfrak{x}$ of* (1) *the part $\mathfrak{x}_1(t)$ tends to* 0 *with $t \to \infty$, and there exists a solution $\mathfrak{z}$ of* (4) *such that*

(8) $\mathfrak{x}_1(t) - \mathfrak{z}_1(t) = 0$ *for* $t = 0$, $\quad \mathfrak{x}_2(t) - \mathfrak{z}_2(t) \to 0$ *for* $t \to \infty$.

Only the statement of the first sentence remains to be proved:

$$\int_0^t U_1(t-\tau)\mathfrak{v}(\tau)\,d\tau \to 0 \text{ for } t \to \infty.$$

Split $\int_0^t$ into $\int_0^{t/2} + \int_{t/2}^t$ and, ϵ being a given positive number, assume that

$$\| U_1(t) \| \leq c_1 \text{ for } t \geq 0, \qquad \int_0^\infty g(\tau)\,d\tau = q;$$

$$\| U_1(t) \| \leq \epsilon \text{ for } t \geq t_\epsilon\ (\geq 0), \qquad \int_{t_\epsilon}^\infty g(\tau)\,d\tau \leq \epsilon.$$

Then as soon as $t \geq 2t_\epsilon$, the absolute value of the first part is less than or equal to $\epsilon \int_0^{t/2} \| \mathfrak{v}(\tau) \|\, d\tau \leq a^* q\epsilon$, that of the second part less than or equal to $c_1 \int_{t/2}^\infty \| \mathfrak{v}(\tau) \|\, d\tau \leq a^* c_1 \epsilon$.

In studying the inverse process, the transition $\mathfrak{z} \to \mathfrak{x}$, we replace (I) by the stronger Lipschitz condition

(I*) $\quad \mathfrak{f}(t, 0) = 0, \qquad \| \mathfrak{f}(t, \mathfrak{x}) - \mathfrak{f}(t, \mathfrak{x}^*) \| \leq \| \mathfrak{x} - \mathfrak{x}^* \| \cdot g(t),$

which again is fulfilled in the linear case $\mathfrak{f}(t, \mathfrak{x}) = G(t)\mathfrak{x}$. For a given $\mathfrak{z}$ the integral equation (5) for $\mathfrak{x}$ may then be solved by successive approximations just as in the classical case $A = 0$. Adding Levinson's hypotheses (II) and (III) we turn at once to the more complicated integral equation (7) and show that it has a unique solution $\mathfrak{x}(t)$ provided

$$\theta = (c_1 + c_2) \int_0^\infty g(\tau)\,d\tau < 1.$$

Proof. Define the successive approximations $\mathfrak{x}_n$ by $\mathfrak{x}_0 = \mathfrak{z}$,

(9) $$\mathfrak{x}_{n+1}(t) = \mathfrak{z}(t) + \int_0^t U_1(t-\tau)\mathfrak{v}_n(\tau)\,d\tau - \int_t^\infty U_2(t-\tau)\mathfrak{v}_n(\tau)\,d\tau$$

$(n = 0, 1, 2, \cdots)$ where $\mathfrak{v}_n(\tau) = \mathfrak{f}(\tau, \mathfrak{x}_n(\tau))$. Since (I*) implies

$$\| \mathfrak{v}_n(t) - \mathfrak{v}_{n-1}(t) \| \leq \| \mathfrak{x}_n(t) - \mathfrak{x}_{n-1}(t) \| \cdot g(t)$$

(even for $n = 0$ if one sets $\mathfrak{x}_{-1}(t) = 0$), the inequality $\| \mathfrak{x}_n(t) - \mathfrak{x}_{n-1}(t) \| \leq a_n$ leads, by means of (9), to

$$\| \mathfrak{x}_{n+1}(t) - \mathfrak{x}_n(t) \| \leq \theta a_n.$$

Hence if $\| \mathfrak{z}(t) \| \leq a$ for $t \geq 0$, then

$$\| \mathfrak{x}_n(t) - \mathfrak{x}_{n-1}(t) \| \leq a\theta^n,$$

and the series $\mathfrak{x}_0 + (\mathfrak{x}_1 - \mathfrak{x}_0) + (\mathfrak{x}_2 - \mathfrak{x}_1) + \cdots$ converges at least as well as the geometric series with the quotient θ. Uniqueness is established by an argument of the same type. I like to arrange it as follows. Suppose a function $\mathfrak{x}(t)$ satisfying the equation (7) is given. Then one finds for the difference $\Delta_n\mathfrak{x} = \mathfrak{x} - \mathfrak{x}_{n-1}$:

$$\Delta_{n+1}\mathfrak{x}(t) = \int_0^t U_1(t-\tau) \cdot \Delta_n\mathfrak{v}(\tau)\,d\tau - \int_t^\infty U_2(t-\tau) \cdot \Delta_n\mathfrak{v}(\tau)\,d\tau$$

where $\Delta_n\mathfrak{v} = \mathfrak{v} - \mathfrak{v}_{n-1}$. If $\| \mathfrak{x}(t) \| \leq a^*$ the last equation yields by induction

$$\| \Delta_n\mathfrak{x}(t) \| \leq a^*\theta^n \qquad (n = 0, 1, 2, \cdots)$$

which proves the uniform convergence of $\mathfrak{x}_n(t)$ toward the given $\mathfrak{x}(t)$. We summarize:

THEOREM. *Under the hypotheses* (I*), (II), (III) *we effect the splitting $U = U_1 + U_2$ described above. Let*

$$\| U_1(t) \| \leq c_1 \textit{ for } t \geq 0, \qquad \| U_2(t) \| \leq c_2 \textit{ for } t \leq 0.$$

If t_0 be so near to infinity that

$$\theta = (c_1 + c_2) \int_{t_0}^\infty g(\tau)\,d\tau < 1$$

then we have established a one-to-one correspondence between the solutions $\mathfrak{x}$ of (1) *and $\mathfrak{z}$ of* (4) *such that*

$$\mathfrak{x}_1(t) - \mathfrak{z}_1(t) = 0 \textit{ for } t = t_0, \qquad \mathfrak{x}_2(t) - \mathfrak{z}_2(t) \to 0 \textit{ for } t \to \infty.$$

INSTITUTE FOR ADVANCED STUDY.

ON THE CONVERGENCE OF SUCCESSIVE APPROXIMATIONS.*

By AUREL WINTNER.

Osgood's criterion for the uniqueness of the initial problem of ordinary differential equations ([3], pp. 344-345) has a dual ([5], p. 283) in which the *local* problem of *uniqueness* becomes replaced by a *non-local* problem of *existence*. In the sequel, another dual of Osgood's theorem will be exhibited, one which replaces the local problem of uniqueness of the solutions by the local problem of the convergence of the process of successive approximations.

Let $f = f(x; t)$, where f, x denote the vectors $(f_1, \cdots, f_m)$, $(x_1, \cdots, x_m)$ with real components f_i, x_i, be a continuous function of the position $(x; t)$ on an $(m+1)$-dimensional region

(1) $$0 \leqq t \leqq a, \qquad |x| \leqq b$$

(the sign of the absolute value refers to Euclidean length). Then the system of m differential equations represented by $x' = f(x; t)$ has, on a sufficiently short, closed t-interval ending at $t = 0$, at least one solution $x = x(t)$ satisfying the initial condition $x(0) = 0$. On the other hand, the mere continuity of $f(x; t)$ does not ensure that the sequence of successive approximations to

(2) $$x' = f(x; t), \qquad x(0) = 0,$$

that is, the sequence of the functions

(3) $$x_{k+1}(t) = \int_0^t f(x_k(s); s)\,ds, \text{ where } x_0(t) = 0,$$

will converge on some (sufficiently short) interval

(4) $$0 \leqq t \leqq c \qquad (c > 0).$$

Along the lines of a counter-example of Picard [4] (which concerns a boundary value problem), this has first been shown by Müller; cf. [2], pp. 35-36. And the only known estimate which does not contain t explicitly and assures the convergence of the sequence (3) on some interval (4) seems to be Lipschitz's

* Received December 17, 1945.

classical condition, requiring the existence of a constant C which satisfies the inequality

(5) $$|f(x^*;t)-f(x^{**};t)| \leqq C\,|x^*-x^{**}|$$

on the region (1).

However, it turns out that, by arguments which, in every respect, are more sophisticated than the "linear" inequalities of Cauchy-Lipschitz-Picard-Lindelöf [1], the condition (5) can be improved so as to become something like a *best* restriction not containing t:

For small, positive values of r, let $\phi(r)$ be a non-negative, monotone, continuous function satisfying

(6) $$\int_{+0} (dr)/\phi(r) = \infty$$

(which implies that

(6 bis) $$\phi(0)=0,$$

if $\phi(0)$ is defined to be $\phi(+0)$). Let $f(x;t)$ be a function which is given and continuous on a region (1) *and is subject there to the inequality*

(7) $$|f(x^*;t)-f(x^{**};t)| \leqq \phi(|x^*-x^{**}|).$$

Then there exists an interval (4) *on which the successive approximations* (3) *to a solution $x=x(t)$ of* (2) *(exist and) are convergent (and converge, uniformly, to a solution).*

Neither of the conditions of smoothness imposed on $\phi(r)$ before (6) will be fully used and, as will be seen at the end of the paper, the first of these conditions is made superfluous by the second: *$\phi(r)$ need not be restricted to be monotone (if it is, e.g., continuous).* The parenthetical amplifications which follow (7) are trivialities, of a general nature, which have nothing to do with the assumption (7). The function f and x can be vectors.

Since (6) is satisfied when

(8) $$\phi(r)/r = \log 1/r,\ \ (\log 1/r)(\log\log 1/r), \cdots,$$

it follows from (7) that the constant C occurring in (5) can be relaxed to any of the functions contained in the set (8). But it is worth noting that ϕ need not be an L-function. And it is just this circumstance that suggests (6) to be the ultimate condition of its kind.

Osgood's theorem, referred to above, states that, on a sufficiently short interval (4), the assumptions (6), (7) prevent the existence of more than

one path $x = x(t)$ satisfying (2) (even if ϕ is not monotone). And this time it is trivial (cf. [5], p. 284) that (6) is the best condition of its kind. Since the crucial step in the proof of the italicized theorem will involve (partly in a recondite connection, namely, via (17) below) applications of Osgood's criterion for uniqueness, one might even expect that, whether (6) and (7) be satisfied or not, *the mere uniqueness of the solution is sufficient for the convergence* of the successive approximations. However, *this was refuted* by Müller's example, quoted above, which he proved to belong to a point $(x; t) = (0; 0)$ of uniqueness. That *also the converse is false,* is shown by the familiar example of the envelope $x = 0$ of the solutions $x = x(t)$ of $x'^3 = x$.

The proof of the theorem proceeds as follows:

Supose first only that $f(x; t)$ is continuous on a region (1). Then (3) defines a sequence

$$(9) \qquad x_1(t), x_2(t), x_3(t), \cdots$$

which need not converge on any interval (4). However, this sequence always is defined on an interval (4), is subject there to the inequalities $|x_k(t)| \leqq b$, and every $x_k(t)$ has there a continuous derivative satisfying $|x_k'(t)| \leqq M$, where M is a constant. Accordingly, (9) is a uniformly bounded sequence of equicontinuous functions on an interval (4). Consequently, (9) contains a subsequence, say

$$(10) \qquad x_h(t), x_i(t), x_j(t), \cdots \qquad (h < i < j < \cdots),$$

which is uniformly convergent on the interval (4). For the sake of brevity, let any such subsequence of (9) be called a selected sequence. In view of the compactness of uniformly bounded, equicontinuous functions, an application of the principle of the excluded middle proves that (9) is uniformly convergent if (and, of course, only if)

$$(11) \qquad y(t) \neq z(t)$$

does not occur for any t, where $y(t)$ and $z(t)$ denote the limit functions of two *unspecified* selected sequences.

If (10) is a selected sequence, then

$$(12) \qquad x_{h+1}(t), x_{i+1}(t), x_{j+1}(t), \cdots$$

must be a selected sequence. In fact, since the functions (9) are uniformly bounded on the interval (8), it is clear from (3) that, if the functions (10) tend to a limit function, say $y(t)$, then the functions (12) must tend to the function

$$y^*(t) = \int_0^t f(y(s);s)\,ds. \tag{13}$$

If the function represented by this integral is identical with the function $y(t)$ occurring beneath the integral sign, then it is, of course, a solution $x = x(t)$ of (2) (though not necessarily the only solution). But it can happen that this is not the case. In fact, this possibility is realized by the Picard-Müller constructions, referred to above.

It will be shown that this possibility cannot arise in the present case. In fact, it will be proved that the assumptions of the theorem imply the existence of an interval (4) on which, without any selection, the difference between two consecutive terms of the sequence (9) tends to 0; that is, on which interval the function

$$\mu(t) = \limsup |w_k(t)|, \qquad (k \to \infty), \tag{14}$$

where

$$w_k(t) = x_{k+1}(t) - x_k(t), \tag{15}$$

vanishes identically. And this will prove the theorem, since the situation is as follows:

If it is known that (15) tends to 0, then, since $y(t)$ and $y^*(t)$ in (13) denote the limit functions of two *consecutive* selected sequences, (10) and (12), these two limit functions must be identical on the interval (4). In other words, (13) becomes

$$y(t) = \int_0^t f(y(s);s)\,ds,$$

which means that $x = y(t)$ is a solution of (2). But (2) cannot have more than one solution, since (6) is precisely Osgood's criterion for uniqueness. Consequently, the limit functions of all selected sequences (10) are the same. In view of the observation made in connection with (11), this implies the convergence of the sequence (9) on the interval (4).

In order to prove the identical vanishing of the function (14), put

$$\Delta g(t) = g(t + \Delta t) - g(t)$$

where $\Delta t > 0$ and $g = \mu, w_k, \cdots$. Then it is clear from (14) that

$$|\Delta\mu(t)| \leqq \limsup |\Delta w_k(t)| \qquad (k \to \infty).$$

But (3) and (15) imply that

$$| \Delta w_{k+1}(t) | \leqq \int_t^{t+\Delta t} | f(x_{k+1}(s); s) - f(x_k(s); s) | \, ds.$$

Furthermore,

$$| f(x_{k+1}(s); s) - f(x_k(s); s) | \leqq \phi(| w_k(s) |),$$

by (7) and (15). And the last three formula lines show that

$$| \Delta\mu(t) | \leqq \limsup \int_t^{t+\Delta t} \phi(| w_k(s) |) ds \qquad (k \to \infty).$$

If $p_1, p_2, \cdots$ are non-negative, measurable functions, then

$$\int_t^{t+\Delta t} \liminf p_k(s) ds \leqq \liminf \int_t^{t+\Delta t} p_k(s) ds \qquad (k \to \infty)$$

(Fatou). And, if C is a sufficiently large constant, this inequality is applicable to the functions

$$p_1 = C - \phi(| w_1 |), \qquad p_2 = C - \phi(| w_2 |), \cdots.$$

In fact, the functions (9) are uniformly bounded; hence, the same is true of the functions (15), and so the continuity of the function $\phi(r)$ assures the existence of a constant C satisfying $\phi(| w_k(t) |) < C$ on (4).

Clearly, the last three formula lines imply that

$$| \Delta\mu(t) | \leqq \int_t^{t+\Delta t} \limsup \phi(| w_k(s) |) ds \qquad (k \to \infty).$$

On the other hand, if t and an $\epsilon > 0$ are fixed, the definition (14) shows that there exists an N having the property that $| w_k | < \mu + \epsilon$ if $k > N$. Since $\phi(r)$ has been assumed to be a non-decreasing function, it follows that $\phi(| w_k |) \leqq \phi(\mu + \epsilon)$ if $k > N$ and so, since $\phi(r)$ is continuous,

$$\limsup \phi(| w_k |) \leqq \phi(\mu) \qquad (k \to \infty).$$

Consequently, by the preceding formula line,

$$| \Delta\mu(t) | \leqq \int_t^{t+\Delta t} \phi(\mu(s)) ds. \tag{16}$$

The inequality (16) for the increments of the non-negative, bounded

function (14) implies that $\mu(t) \to 0$ as $t \to 0$ and that, in the sense of a parametrised Lebesgue-Stieltjes integration,

$$(17) \qquad |\, d\mu \,|/\phi(\mu) \leqq dt,$$

since $\phi(t) \geqq 0$. If this differential inequality is integrated between $t = \epsilon\ (> 0)$ and a fixed $t = t_0\ (> \epsilon)$, it is seen that, if the function $\mu(t)$ did not vanish identically on an interval (4), the assumption (6) would lead to the following contradiction:

$$(18) \qquad + \infty \leqq t_0 - \lim_{\epsilon \to 0} \epsilon.$$

However, this conclusion is legitimate only if the mapping of the measures which is defined by the transformation $t \to \mu$ is sufficiently smooth; for instance, if $\mu(t)$ is continuous. In fact, if $\mu(t)$ is continuous, it must attain every μ-value contained between $\mu(+0) = 0$ and $\mu(t_0)$, and so the argument leading to the contradiction (18) becomes justified. But it turns out that $\mu(t)$ *is* continuous. In order to see this, it is sufficient to observe that, according to (3), (15) and (14), the function $\mu(t)$ is the upper limit of a sequence of uniformly bounded, equicontinuous functions on the interval (4). In fact, as verified by Montel in § 20-§ 21 of his Thèse (1907), the upper limit of any such sequence is always continuous.

This completes the proof of the theorem.

That the restriction of monotony, imposed on $\phi(r)$ before (6), is superfluous, follows if the consideration of the functions (15) themselves is replaced by that of their "best monotone majorant sequence," that is, of the functions

$$\operatorname*{Max}_{0 \leqq s \leqq t} |\, x_{k+1}(s) - x_k(s) \,|.$$

Corresponding to the circumstance that the above proof depends, via the criterion (11), on a "suppose the contrary" argument, there results no explicit estimate of the deviation of the solution $x(t)$ from the approximations (3). I do not know whether this is due to the *unrestricted* character of a (monotone) function ϕ allowed in (6), (7) or there can exist *general* theorems on (monotone) transformations, providing, as in the particular case (5), a direct approach.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] E. Lindelöf, "Sur l'application des méthodes d'approximations successives à l'étude des integrales réelles des équations différentielles ordinaires," *Journal de Mathématiques Pures et Appliquées*, ser. 4, vol. 10 (1894), pp. 117-128.

[2] M. Müller, "Über die Eindeutigkeit der Integrale eines Systems gewöhnlicher Differentialgleichungen und die Konvergenz einer Gattung von Verfahren zur Approximation dieser Integrale," *Sitzungsberichte der Heidelberger Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse*, Jahrgang 1927, no. 9, pp. 1-38.

[3] W. F. Osgood, "Beweis der Existenz einer Lösung der Differentialgleichung $dy/dx = f(x, y)$ ohne Hinzunahme der Cauchy-Lipschitzschen Bedingung," *Monatshefte der Mathematik und Physik*, vol. 9 (1898), pp. 331-345.

[4] É. Picard, "Sur un exemple curieux d'approximations successives divergentes, *Comptes Rendus*, vol. 118 (1894), pp. 899-902.

[5] A. Wintner, "The non-local existence problem of ordinary differential equations," *American Journal of Mathematics*, vol. 67 (1945), pp. 277-284.

A REMARK ON THE MAPPING OF MULTIPLY-CONNECTED DOMAINS.*

By STEFAN BERGMAN.

1. Introduction. The fundamental theorem of the theory of conformal mapping states that the universal covering surface of a multiply-connected domain can be mapped by a conformal and one-to-one transformation on the unit circle.

Using this result it is possible to prove that every multiply connected domain satisfying certain conditions can be mapped into characteristic domains of various types.

A characteristic domain is a domain which possesses certain geometric characterising properties,[1] e. g. a ring domain with a number of slits along arcs of concentric circles. In the following, we shall refer to *this* domain as a characteristic domain $\mathfrak{C}$.

To every multiply connected domain there exist a finite number of characteristic domains $\mathfrak{C}$. Thus knowing the mapping functions of two multiply connected domains, B_k, $k = 1, 2$, into the corresponding characteristic domains, we can decide whether the domains B_k can be mapped into one another by a conformal and one-to-one transformation.

In the present note, using certain orthogonal functions, a formula for the function $w(z)$, which maps a multiply connected domain into the characteristic domain $\mathfrak{C}$ will be given.

The interest of the expression derived below lies not merely in that it can be used as a tool for the actual computing of the required mapping function, and thus make it possible actually to determine when two multiply connected domains can be mapped into each other, but that these results admit the following applications:

Using the theory of orthogonal functions,[2] introducing the kernel func-

* Received May 4, 1945.

[1] Various types of characteristic domains were introduced by Courant [3], by Hilbert, by Koebe [5], and by Schottky [8]. Numbers in brackets refer to the bibliography.

[2] Using orthogonal functions of a complex variable the theory of multiply connected domains has been treated in a manner somewhat different from the classical approach. Instead of Poincaré's metric being introduced on the universal covering surface, a new

tion and with its aid an invariant metric, it is possible to develop a method of attack which, without using Riemann's theorem, leads to various results in the case of simply connected domains, (e. g. in the theory of distortion). See [2], [12]. Certain of these results can now be extended to the case of multiply connected domains.

One of the advantages of this approach is that it can be generalized to the case of mappings by n analytic functions of n complex variables. We do not discuss these generalizations in the present note.

2. A representation of the mapping functions of the universal covering surface into the unit circle. Let B be a bounded $(p+1)$-connected domain whose boundary, b, consists of $p+1$ regular closed curves, b_ν $(\nu = 0, 1, \cdots, p;\ p \geqq 1)$; we may suppose b_0 to be the exterior curve. It will now be shown how to determine the function which maps the universal covering surface G of the domain B into the unit circle. Let α_ν $(\nu = 1, 2, \cdots, p)$ be exterior points of the domain such that α_ν lies in the "hole" bounded by b_ν. Introduce the auxiliary domain P which consists of the entire plane, out of which have been cut the points $\alpha_1, \cdots, \alpha_p$ and the point at infinity. Let M be the universal covering surface of P; then each of the points α_ν and ∞ becomes a branch point of infinite order of M. See e. g., [7]. Clearly the universal covering surface G is a part of M.

It is well-known that there exists a function $z(\lambda)$—modular function—which maps the upper half of the λ-plane on the universal covering surface M. The inverse of this transformation, $\lambda(z)$, maps any conveniently cut sheet of M on a fundamental domain F in the λ-plane. The fundamental domain F is a simply-connected domain bounded by $2p$ semicircles in the upper half-plane which are tangent to each other and orthogonal to the real axis of the λ-plane. Thus, to each sheet of M there will correspond a fundamental domain F in the λ-plane, and it is evident that the same transformation will map the sheets of G into simply-connected portions R of the corresponding fundamental domains F. The sheets of M, and therefore those of G, may be enumerated.

metric (which in general has a non-constant curvature) was defined for multiply connected domains, [1], [2], [9]. Further another kind of characteristic domains (so called representative and minimal domains) has been considered.

Zarankiewicz has studied the above metric in a detailed manner for the case of doubly connected domains [9], [10], and Kufareff has determined the geometric shape of minimal domains [6] in this case. Greenstone [12] has established certain interesting properties of the above metric in the general case of multiply connected domains. The image in the domain B of the Cartesian coordinate frame of the representative domain of B is termed the "representative coordinate frame" of B. In [11] and papers cited there Fuchs has studied properties of representative coordinate systems.

Let G_k be the sheet of G which lies on the sheet M_k of M. Suppose that the sheets G_k have been so joined and enumerated that $\sum_{k=1}^{n} G_k$ is simply connected for every n: That this may always be done is obvious.[3]

The function $\lambda(z)$ maps G_1 into the domain R_1 which lies in the upper half of the λ-plane, as already described. Let λ_0 be an interior point of R_1; and let $w_1(\lambda)$ be the function which maps R_1 into the unit circle, taking λ_0 into the origin. $\sum_{k=1}^{n} G_k$ is a simply-connected domain which is mapped by $\lambda(z)$ into $\sum_{k=1}^{n} R_k$, which is also a simply-connected domain containing λ_0 in its interior. Therefore there exists a function $w_n(\lambda)$ which maps $\sum_{k=1}^{n} R_k$ into the unit circle, taking λ_0 into the origin. $G = \sum_{k=1}^{\infty} G_k$ is the kernel domain (in the sense of Carathéodory) of the sequence of domains $\sum_{k=1}^{n} G_k$, and therefore, by the theorem of Carathéodory [3], $\lim_{n\to\infty} w_n[\lambda(z)]$ exists and maps G into the unit circle. Thus,

$$w(z) = \lim_{\nu\to\infty} w_\nu[\lambda(z)] \tag{2.1}$$

represents the required mapping function which maps G into the unit circle.

The function (2.1) can immediately be expressed in terms of orthogonal functions.

Indeed, denote by

$$\{\Psi_k^{(\nu)}\}, \qquad (k = 1, 2, 3, 4, \cdots) \tag{2.2}$$

the set of orthogonal functions which we obtain by orthogonalizing the powers λ^k $(k = 0, 1, 2, \cdots)$, with respect to $\sum_{n=1}^{\nu} G_n$. Then the function $w_\nu(\lambda)$ which maps $\sum_{n=1}^{\nu} G_n$ into the unit circle can be written in the form

$$w_\nu(\lambda) = \frac{\pi^{\frac{1}{2}} \int_0^\lambda \sum_{k=1}^{\infty} \Psi_k^{(\nu)}(\lambda) \overline{\Psi_k^{(\nu)}(\lambda_0)}\, d\lambda}{\left[\sum_{k=1}^{\infty} | \Psi_k^{(\nu)}(\lambda_0) |^2\right]^{\frac{1}{2}}}. \tag{2.3}$$

The formulae (2.1), (2.3) yield the mapping of the universal covering surface into the unit circle in the w-plane. The functions which map B into characteristic domains can be expressed in a well-known manner, in the form of an infinite series in terms of $w(z)$. See e. g. [5].

---

[3] If any two simply connected domains, say G_1 and G_2, have a simple open curve as their only common boundary, then the combined domain $G_1 + G_2$ is also simply connected. This is exactly the case we have here, because G_1 and G_2 are joined along one

On the other hand the inverse of the modular function appears in (2.1), (2.3); further (2.1) is expressed not directly in terms of orthogonal functions, but as a limit of a series development of this kind. This fact often causes certain difficulties, e. g. in actual computation of the mapping function, in the distortion theory, etc. In the next section we shall derive an expression for the mapping function $w(z)$ of B into the characteristic domain $\mathfrak{C}$. The formula obtained avoids the above disadvantages; we express $w(z)$ in terms of certain functions which we obtain by orthogonalizing the powers z^n $(n = 0, 1, 2, \cdots)$, $(z - \alpha_\nu)^{-h}$ $(\nu = 1, 2, \cdots, p,\ h = 1, 2, \cdots)$. Here α_ν are the points introduced above. On the other hand in the formula certain unknown constants appear. These constants can be determined by certain additional considerations, which are however not discussed in detail in the present paper.

3. The determination of an analytic function from boundary values of its real part. Let B be the domain introduced in **2**, whose boundary b consists of $p + 1$ regular closed curves b_ν $(\nu = 0, 1, 2, \cdots, p,\ p \geqq 1)$.

In the following we shall denote by $\{\phi'_k(z)\}$ $(k = 1, 2, 3, \cdots)$, a complete system of orthogonal functions, i. e. a system for which

$$\text{(3.1)} \qquad \iint_B \phi'_k(z)\overline{\phi'_s(z)}\,d\omega = 1 \text{ for } k = s,\ d\omega = dxdy, \qquad = 0 \text{ for } k \neq s,$$

and such that for every analytic function $g(z)$, $\iint_B |g(z)|^2 d\omega < \infty$,

$$\text{(3.2)} \qquad \iint_B |g(z)|^2\, d\omega = \sum_{k=1}^{\infty} \left| \iint_B g(z)\overline{\phi'_k(z)}\,d\omega \right|^2.$$

Let a be an interior point of B. We shall consider in the following the functions

$$\text{(3.3)} \qquad \phi_k(z) = \psi_k(x, y) + i\theta_k(x, y), \qquad \phi_k(a) = 0.$$

Remark. Since B is a multiply connected domain the functions $\phi_k(z)$ are multivalued. See examples (3.10) and (3.11). In order to avoid any ambiguity, we introduce the simply connected domain $\tilde{B}$ obtained from B by p conveniently chosen cuts. The functions $\phi_k(z)$ are single valued in $\tilde{B}$.

THEOREM 3.1. *Let $f(z) = f_1(x, y) + if_2(x, y)$ be a function which is regular in B, whose derivative $f'(z)$ as well as the real part f_1 are single valued*

of the simple curves along which B has been cut. By induction it follows immediately that this holds for any number of simply connected domains.

in B. $f(z)$ is assumed to map B into a domain with a finite area, i. e. it is assumed that

$$\iint_B |f'(z)|^2 d\omega < \infty. \tag{3.4}$$

Then $f(z)$ can be represented in $\tilde{B}$ in the form

$$f(z) = i\sum_{k=1}^{\infty} \phi_k(z) \int_b f_1(\xi,\eta)\, d\overline{\phi_k(\xi+i\eta)} + f(a), \qquad z \in \tilde{B}. \tag{3.5}$$

Proof. By the theorem on the development in the series of orthogonal functions, see [2] p. 47, we can write

$$f'(z) = \sum_{k=1}^{\infty} \phi'_k(z) \iint_B f'(Z)\overline{\phi'_k(Z)}\, d\Omega, \quad d\Omega = dXdY, \quad Z = X + iY. \tag{3.6}$$

Now

$$\begin{aligned}
\iint_B f'(Z)\overline{\phi'_k(Z)}\, d\Omega &= \iint_{\tilde{B}} f'(Z)\overline{\phi'_k(Z)}\, d\Omega \\
&= \iint_{\tilde{B}} [f_{1,X}\psi_{k,X} + f_{2,X}\theta_{k,X}]\, d\Omega + i \iint_{\tilde{B}} [f_{2,X}\psi_{k,X} - f_{1,X}\theta_{k,X}]\, d\Omega \\
&= \iint_{\tilde{B}} [f_{1,X}\psi_{k,X} + f_{1,Y}\psi_{k,Y}]\, d\Omega - i \iint_{\tilde{B}} [f_{1,X}\theta_{k,X} + f_{1,Y}\theta_{k,Y}]\, d\Omega \\
&= \int_b f_1[-\frac{\partial\psi_k}{\partial n}\, ds + i\frac{\partial\theta_k}{\partial n}\, ds] \\
&\quad + \sum_{\nu=1}^{2p} \int_{e_\nu} f_1[-\frac{\partial\psi_k}{\partial n}\, ds + i\frac{\partial\theta_k}{\partial n}\, ds],
\end{aligned} \tag{3.7}$$

$f_{1,X} \equiv (\partial f_1/\partial X)$, $\psi_{k,X} \equiv (\partial\psi_k/\partial X)$ etc.

where $e_{2\nu}$, $e_{2\nu-1}$ $(\nu = 1, 2, \cdots, p)$, are the edges of crosscuts which dissect B into a simply connected domain $\tilde{B}$. dn and ds denote an element of the interior normal and the line element, respectively. Since f_1, $\partial\psi_k/\partial n$ and $\partial\theta_k/\partial n$ are single valued in B, the integrals over e_ν vanish and we find that (3.7) equals

$$\int_b f_1[d\theta_k + id\psi_k] = i\int_b f_1(\xi,\eta)\, d\overline{\phi_k(\zeta)}. \tag{3.8}$$

A system of orthogonal functions satisfying the above conditions can be obtained by orthogonalizing the system

$$1,\ z^k,\ (z - \alpha_\mu)^{-k}, \qquad (\mu = 1, \cdots, p,\ k = 1, 2, \cdots) \tag{3.6}$$

where α_μ is a point which lies in the "hole" bounded by b_μ. See [2], p. 57. In this case the functions ϕ_k are of the following form:

$$(3.10)\qquad \begin{aligned}
\phi_1 &= \beta_{1,0} + \beta_{1,1}z, \\
\phi_2 &= \beta_{2,0} + \beta_{2,1}z + \beta_{2,2}z^2, \\
\phi_{3+\mu} &= \beta_{3+\mu,0} + \beta_{3+\mu,1}z + \beta_{3+\mu,2}z^2 + \sum_{h=0}^{\mu} \beta_{3+\mu,3+h} \log(z-\alpha_h), \\
&\qquad\qquad (\mu = 0, 1, \cdots, p), \\
\phi_{4+p} &= \beta_{4+p,0} + \beta_{4+p,1}z + \beta_{4+p,2}z^2 + \sum_{h=0}^{p} \beta_{4+p,3+h} \log(z-\alpha_h) + \beta_{4+p,4+p}z^2, \\
\phi_{5+p+s} &= \beta_{5+p+s,0} + \beta_{5+p+s,1}z + \beta_{5+p+s,2}z^2 \\
&\quad + \sum_{h=0}^{p} \beta_{5+p+s,3+h} \log(z-\alpha_h) + \beta_{5+p+s,4+p}z^3 \\
&\quad + \sum_{g=0}^{s} \beta_{5+p+s,5+p+g}(z-\alpha_g)^{-1}, \qquad (s = 0, 1, \cdots, p),
\end{aligned}$$

Example. If $B = \mathrm{E}[r < |z| < 1]$ then $p = 1$. We choose $\alpha_1 = 0$.

$$(3.11)\qquad \begin{aligned}
\phi_1 &= (z-a)/[\pi(1-r^2)]^{\frac{1}{2}}, \qquad \phi_3 = (\log z - \log a)(-2\pi \log r)^{-\frac{1}{2}}, \\
\phi_{2n-2} &= (z^n - a^n)[\pi n(1-r^{2n})]^{-\frac{1}{2}}, \qquad (n = 2, 3, \cdots), \\
\phi_{2n-1} &= (a^{1-n} - z^{1-n})[\pi(1-n)(1-r^{-2(n-1)})]^{-\frac{1}{2}}, \quad (n = 3, 4, \cdots).
\end{aligned}$$

4. The representation of integrals of the first kind in terms of orthogonal functions. The domain B being given, and two boundary curves, b_1 and b_0, being prescribed there exists a function $w(z)$ which maps B into $\mathfrak{C}$ i.e. the circular ring.[4] $\mathrm{E}[\exp a_1 < |z| < 1]$ with $(p-1)$ slits along arcs $\mathrm{E}[\alpha_{\nu 1} < \arc z < \alpha_{\nu 2}, |z| = \exp a_\nu]$ $(\nu = 2, \cdots, p)$, $a_1 < a_2 \leqq a_3 \cdots \leqq a_{p-1} < 1$. See [5]. Two prescribed components, b_0 and b_1 of the boundary b of B are transformed into the interior and exterior circles, respectively.

In this section we shall express the functions

$$(4.1)\qquad s(z) = [\log w(z)]$$

in terms of functions $\phi_k(z)$, introduced in (3.9).

The function $s(z)$ possesses the following properties:

I. (4.2) $\mathrm{Re}[s(z)] = a_\nu$ on b_ν, $(\nu = 0, 1, \cdots, p-1, a_0 = 0)$.

II. The function $s(z)$ maps the conveniently cut domain B into the rectangle

$$\mathrm{E}[a_1 < \xi < 0, 0 < \eta < 2\pi]$$

[4] By E [...] we denote a set of points whose coordinates satisfy the inequalities indicated in brackets.

with $(p-1)$ slits along the straight lines $\xi = a_\nu$ $(\nu = 1, 2, \cdots, p-1)$. (These slits are not necessarily connected.)

III. The derivative $s'(z)$ as well as $\operatorname{Re}[s(z)]$ are single valued functions in B.

THEOREM 4.1. *The function $s(z)$ can be represented in B in the form*

$$s(z) = i \sum_{k=1}^{\infty} \phi_k(z) \sum_{\mu=1}^{p} a_\mu \Delta_\mu \overline{\phi_k(z)} + s(a) \tag{4.3}$$

where $\Delta_\mu \overline{\phi_k(z)}$ denotes the increase of the function $\overline{\phi_k(z)}$ as we move around b_μ $(\mu = 0, 1, 2, \cdots, p-1)$.

Proof. Since $s(z)$ satisfies all the hypotheses of Theorem 3.1, we can write it in the form (3.5). In the case under consideration we have

$$\int_b f_1 d\bar{\phi}_k = \sum_{\mu=0}^{p} a_\mu \int_{b_\mu} d\bar{\phi}_k = \sum_{\mu=1}^{p} a_\mu \Delta_\mu \bar{\phi}_k \tag{4.4}$$

from which (4.3) follows. (Note that $a_0 = 0$).

Example. Suppose $B = \mathrm{E}[r < |z| < 1]$ and $a_1 = 0$. According to (3.11) all $\Delta\phi_k = 0$, except for $k = 3$, for which we have

$$\Delta\phi_3 = (-2\pi \log r)^{-\frac{1}{2}} \cdot 2\pi i,$$

and therefore

$$s(z) = i \log r\, (-2\pi \log r)^{-1} \cdot 2\pi i (\log z - \log a) + \log a = \log z.$$

Remark. Except $\log(z - a_\mu)$ all expressions in (3.7) do not change if we move around b_μ. Hence

$$\Delta_\mu \phi_k = -2\pi i \beta_{k,\mu+3} \tag{4.5}$$

where $\beta_{k,\mu}$ are the quantities introduced in (3.7).

Remark. p unknown constants a_μ appear in the expression (4.3). We note that the theory of orthogonal functions provides the means for determining them. The procedure is based on the fact that since $\operatorname{Re}[s(z)] = a_\mu$ on b_μ $(\mu = 1, 2, \cdots, p)$,

$$\operatorname{Re}[\lim \int_{z_1}^{z_2} s(z)\, dz] = a_{\mu_2} - a_{\mu_1}.$$

Here z_1 and z_2 are points which lie in B and are such that $\lim z_k \in b_{\mu_k}$ $(k = 1, 2)$. We omit a more detailed determination of the constants here.

5. An application to the theory of distortion. As is indicated in more detail in [2] § 9 the formulae expressing functions in terms of orthogonal functions often yield inequalities for the functions under consideration. As an example of the application of these methods, an inequality for $s'(z) = d[\log w(z)]/dz$ will be derived in this section.

THEOREM 5. 1. *Let c_ν, $c_\nu \in B$, be p closed simple curves with the following properties:*

1. *The distance of every point of c_ν from the boundary b of B is greater than δ_1.*

2. *By a one-to-one and continuous transformation each c_ν can be reduced to b_ν without cutting the remaining b_μ, $\mu \neq \nu$.*

Let $w(z)$ be the function which maps B into the characteristic domain $\mathfrak{C}$ i. e. $\mathrm{E}[\exp a_1 < |z| < 1]$ with $p-1$ slits along arcs situated on circles $|z| = \exp a_\nu$ $(\nu = 2, 3, \cdots, p)$.

Further let z be a point of B, whose distance from the boundary is larger than δ_2. Then

$$\left|\frac{d \log w(z)}{dz}\right| \leqq \sum_{\nu=1}^{p} \frac{|a_\nu| \, l(c_\nu)}{\pi \delta_1 \delta_2} \tag{5. 1}$$

where $l(c_\nu)$ denotes the length of c_ν.

Proof. According to (3. 5) and (4. 4)

$$\frac{d \log w(z)}{dz} = \frac{ds(z)}{dz} = i \sum_{k=1}^{\infty} \phi'_k(z) \sum_{\nu=0}^{p} a_\nu \int_{b_\nu} \overline{\frac{d\phi_k(\zeta)}{d\zeta}} \, d\bar{\zeta}. \tag{5. 2}$$

From 2. it follows that the integration curves b_ν can be replaced by the c_ν. Thus (5. 2) equals [5]

$$\begin{aligned} (5.3) \qquad & i \sum_{k=1}^{\infty} \phi'_k(z) \sum_{\nu=0}^{p} a_\nu \int_{c_\nu} \overline{\frac{d\phi_k(\zeta)}{d\zeta}} \, d\bar{\zeta} \\ &= i \sum_{\nu=1}^{p} a_\nu \int_{c_\nu} \sum_{k=1}^{\infty} \phi'_k(z) \overline{\phi'_k(\zeta)} \, d\bar{\zeta} \\ &= i \sum_{\nu=1}^{p} a_\nu \int_{c_\nu} \mathrm{K}_B(z, \bar{\zeta}) \, d\bar{\zeta}. \end{aligned}$$

Here $\mathrm{K}_B(z, \bar{\zeta})$ denotes the kernel function of B. See [2], § 7. From hypothesis 1, the assumption concerning the location of z and (7. 18) p. 47 of [2],

[5] By the theorem on p. 47 of [2] the series $\sum_{\nu=1}^{\infty} |\phi'_\nu(z) \phi'_\nu(\zeta)|$ converges uniformly in every closed domain which lies in B.

$$K_B(z,\bar{\zeta}) \leqq \left[\sum_{\nu=1}^{\infty} |\phi'_\nu(z)|^2 \cdot \sum_{\nu=1}^{\infty} |\phi'_\nu(\zeta)|^2\right]^{\frac{1}{2}} \leqq \pi^{-1}\delta_1^{-1}\delta_2^{-1},$$

which yields the inequality (5. 1).

We note that using other inequalities for $K_B(z,\bar{\zeta})$ and for its derivatives (see [2] § 9) one can obtain other inequalities similar to (5. 1).

BROWN UNIVERSITY.

BIBLIOGRAPHY.

1. Stefan Bergman, "Über die Kernfunktion eines Bereiches und ihr Verhalten am Rande," *Journ. für die reine u. ang. Math.*, vol. 169 (1933), pp. 1-42 and vol. 172 (1934), pp. 89-128.

2. ———, *Partial differential equations, Advanced Topics* (*Conformal Mapping of multiply connected domains*) Publication of Brown University, Providence, R. I., 1941.

3. R. Courant, "Plateau's problem and Dirichlet's principle," *Annals of Mathematics*, vol. 38 (1937), pp. 679-724.

4. C. Carathéodory, "Untersuchungen über die konforme Abbildungen von festen und veränderlichen Gebieten," *Mathematische Annalen*, vol. 72 (1912), pp. 107-144.

5. Paul Koebe, "Abhandlungen zur Theorie der konformen Abbildung II, IV," *Acta Mathematica*, vol. 40 (1915/16), pp. 251-290 and vol. 41 (1916), pp. 305-344.

6. Paul Kufareff, "Über das zweifach zusammenhängende Minimalgebiet," *Bull. Inst. Math. et mec. Université de Tomsk*, vol. 1 (1935-1937), pp. 228-236

7. R. Nevanlinna, *Eindeutige analytische Funktionen*, Berlin, 1936.

8. F. Schottky, "Über die conforme Abbildung mehrfach zusammenhängender ebener Flächen," *Journ. für die reine u. ang. Math.*, vol. 83 (1877), pp. 300-351.

9. K. Zarankiewicz, "Sur la représentation conforme d'un domaine doublement connexe sur un anneau circulaire," *C. R. Acad. Sc.*, Paris, vol. 198 (1934), pp. 1347-1349.

10. ———, "Über ein numerisches Verfahren zur konformen Abbildung zweifach zusammenhängen Gebiete," *Zeit. f. ang. Math. u. Mech.*, vol. 14 (1934), pp. 97-104.

11. B. Fuchs, "Sur la fonction minimale d'un domaine, I," *Mat. Sbornik* (*Recueil Mathem.*), N. S. vol. 16 (58) (1945), pp. 21-38.

12. L. Greenstone, "Mapping of simply and multiply connected domains by analytic functions," to appear.

ON A PROBLEM OF RAMANUJAN.*

By Arnold E. Ross.

1. It was apparently known to Diophantus and first proven by Lagrange (8) that the form $x^2 + y^2 + z^2 + u^2$ represents all positive integers. Examples of other integral forms

$$\Phi = ax^2 + by^2 + cz^2 + du^2 \tag{1.1}$$

which represent all positive integers, were first obtained by Jacobi (5), Liouville (9), and Pepin (13). Ramanujan (14) proved that there are only 54 sets of positive integers a, b, c, d such that (1.1) represents all positive integers. Dickson (2) called such forms *universal.* Universal quaternary quadratic forms with cross products were studied by Dickson (2) and Morrow (11).

In the above mentioned paper Ramanujan proposed another problem, *viz.*, the problem of determining the conditions under which positive quadratic forms (1.1) represent all except a finite number of integers. Kloosterman (7), employing the methods of Hardy-Littlewood succeeded, save for a finite number of exceptions, in solving that problem.

It is natural to ask Ramanujan's question concerning general positive quaternary quadratic forms. Should Tartakowsky's theorem (19) concerning the representation of large integers by positive quadratic forms in $n \geqq 5$ variables hold also for $n = 4$, then one would expect the answer to that question to be found as an elementary corollary of this theorem and to be expressed in terms of the generic characters of quadratic forms. It is of interest to note that, although Tartakowsky's theorem does not carry over unconditionally to forms in four variables, still for forms of odd determinants and certain orders of even determinants, the answer to Ramanujan's question may be obtained as an elementary extension of the results of Kloosterman and some other elementary considerations, and that, moreover, save for a finite number (of classes) of exceptions the conditions are given in terms of the generic characters. The results here obtained suggest conditions which the generic characters of a genus of quaternary forms should fulfil in order that all forms of that genus should represent the same large integers.

The method employed may be summarized as follows: Through the use of the canonical form of Section 3, the problem of the representation of in-

* Received August 3, 1942 and June 14, 1945.

tegers by the original form is reduced to that of the representation of integers by a certain quadratic form without the cross products (Section **4.1**). Kloosterman's conditions (7) applied to this last form (Section **5**) yield a set of generic character conditions which assure the representation of all large integers by the original form. Upon closer examination of these conditions one notices (Sections **6** and **7**) that some of these are necessary but that the failure of the remaining merely implies that a form represents *all large* integers only if it represents *all* integers or *all even* integers. In view of the results in Section **2**, the determinants of such forms do not exceed a fixed number. Thus, outside of forms in Section **5**, there is only a finite number of classes of forms representing all large integers. A study of some of these classes (Section **8**) yields interesting examples of representation of integers by positive quaternary forms in a fixed genus.

2. An upper bound for determinants of classic universal positive quaternary quadratic forms. Among the 54 universal forms of type (1.1), the form $x^2+2y^2+4z^2+14u^2$ has [1] the largest determinant 112. A simple extension of Ramanujan's argument yields [2] the more general and quite useful

THEOREM 2. *The determinant of every classic universal positive quaternary quadratic form is* $\leqq 112$.

We write

$$\Phi_1(x) = x'Ax = \sum_{i,j=1}^{4} a_{ij}x_ix_j \tag{2.01}$$

where a_{ij} are integers. If $\Phi_1(x)$ represents all positive integers, it represents 1 properly, and hence is equivalent to a form of type (2.01) with $a_{11}=1$ and $a_{1j}=0$ for $j=2,3,4$. Thus

$$\Phi_1(x) \sim \Phi_2(y) = y_1^2 + \phi_2(y_2, y_3, y_4) = y_1^2 + \sum_{i,j=2}^{4} b_{ij}y_iy_j.$$

In order that Φ_1 and, hence, Φ_2 should represent all positive integers, the minimum a of ϕ_2 must be $\leqq 2$. For otherwise Φ_2, and therefore also Φ_1, would not represent 2. Since the minimum a is represented properly by ϕ_2

$$\phi_2 \sim \phi_3 = az_2^2 + bz_3^2 + cz_4^2 + 2rz_3z_4 + 2sz_2z_4 + 2tz_2z_3, \tag{2.02}$$

where

$$0 \leqq s < a \quad \text{and} \quad 0 \leqq t < a, \qquad a = 1 \text{ or } 2, \tag{2.03}$$

and hence

$$\Phi_1 \sim \Phi_3 = z_1^2 + \phi_3(z_2, z_3, z_4). \tag{2.04}$$

---

[1] Cf. Dickson (4), p. 115.

[2] Ross (15), Theorem 8.

2.1. We let, first, $a = 1$. Then, in view of (2.02)-(2.04),

$$(2.11) \quad \Phi_3 = z_1^2 + z_2^2 + bz_3^2 + cz_4^2 + 2rz_3z_4 = z_1^2 + z_2^2 + \psi_3(z_3, z_4).$$

In order that Φ_3 should represent 3, the minimum M of ψ_3 should be ≤ 3. Then

$$(2.12) \quad \psi_3 \sim \psi_4 = Mu_3^2 + 2Nu_3u_4 + Lu_4^2,$$

$$(2.13) \quad -(M/2) < N \leq M/2, \qquad M = 1, 2, \text{ or } 3,$$

and

$$(2.14) \quad \Phi_1 \sim \Phi_4 = u_1^2 + u_2^2 + \psi_4.$$

The form Φ_1 would represent all integers only if $M\Phi_1$ should represent all multiples of M. But in view of (2.12)-(2.13)

$$M\Phi_1 \sim M\Phi_4 = Mu_1^2 + Mu_2^2 + (Mu_3 + Nu_4)^2 + Du_4^2$$

where $D = ML - N^2$ is the determinant of Φ_1. Thus in order that $M\Phi_1$ should represent all multiples of M, D must not exceed the smallest multiple of M not represented by

$$f_M(u_1, u_2, U_3) = Mu_1^2 + Mu_2^2 + U_3^2.$$

If $M = 1$, $f_1 = u_1^2 + u_2^2 + U_3^2 \neq 7$ and hence $D \leq 7$.
If $M = 2$, $f_2 = 2u_1^2 + 2u_2^2 + U_3^2 \neq 28$ and therefore $D \leq 28$.
If $M = 3$, $f_3 = 3u_1^2 + 3u_2^2 + U_3^2 \neq 18$ and therefore $D \leq 18$.

Thus, in case $a = 1$, there is no universal form Φ_1 of determinant > 28

2.2. Next, let $a = 2$. In order that Φ_1 should be universal $2\Phi_1$ should represent all even integers. But in view of (2.02) and (2.04),

$$2\Phi_1 \sim 2\Phi_3 = 2z_1^2 + (2z_2 + tz_3 + sz_4)^2 + \psi_3(z_3, z_4)$$

where

$$\psi_3(z_3, z_4) = (ab - t^2)z_3^2 + 2(ar - st)z_3z_4 + (ac - s^2)z_4^2.$$

Since $2z_1^2 + Z_2^2 \neq 10$, the minimum M of $\psi_3(z_3, z_4)$ is ≤ 10. Also,

$$\psi_3 \sim \psi_4 = Mu_3^2 + 2Nu_3u_4 + Lu_4^2$$
$$2\Phi_1 \sim 2\Phi_4 = 2u_1^2 + (2u_2 + t_1u_3 + s_1u_4)^2 + \psi_4(u_3, u_4)$$

and

$$2M\Phi_1 \sim 2M\Phi_4 = 2Mu_1^2 + M(2u_2 + t_1u_3 + s_1u_4)^2 + (Mu_3 + Nu_4)^2 + (ML - N^2)u_4^2.$$

In order that Φ_1 should be universal $2M\Phi_1$ should represent all multiples of $2M$. Hence $ML - N^2$ does not exceed the smallest multiple of $2M$ not represented by

$$f_M(u_1, U_2, U_3) = 2Mu_1^2 + MU_2^2 + U_3^2.$$

Since by a well known theorem on determinants the determinant D of Φ_1 is equal to $\frac{1}{2}(ML - N^2)$, we have the following results which we state in a schematic form:

M	$f_M = 2Mu_1^2 + MU_2^2 + U_3^2 \neq 2Mk$	$ML - N^2 \leq 2Mk$	$D \leq$
1	$2u_1^2 + U_2^2 + U_3^2 \neq 14$	≤ 14	≤ 7
2	$4u_1^2 + 2U_2^2 + U_3^2 \neq 56$	≤ 56	≤ 28
3	$6u_1^2 + 3U_2^2 + U_3^2 \neq 30$	≤ 30	≤ 15
4	$8u_1^2 + 4U_2^2 + U_3^2 \neq 56$	≤ 56	≤ 28
5	$10u_1^2 + 5U_2^2 + U_3^2 \neq 50$	≤ 50	≤ 25
6	$12u_1^2 + 6U_2^2 + U_3^2 \neq 120$	≤ 120	≤ 60
7	$14u_1^2 + 7U_2^2 + U_3^2 \neq 98$	≤ 98	≤ 48
8	$16u_1^2 + 8U_2^2 + U_3^2 \neq 224$	≤ 224	≤ 112
9	$18u_1^2 + 9U_2^2 + U_3^2 \neq 126$	≤ 126	≤ 63
10	$20u_1^2 + 10U_2^2 + U_3^2 \neq 200$	≤ 200	≤ 100

Thus in every case the determinant D does not exceed 112.

2.3. We have just seen that there is but a finite number of classes of classic universal positive quaternary quadratic forms. We inquire next whether the same would be true of forms which, although not universal, do nevertheless represent all even integers. We show that

THEOREM 2.3. *The determinants of classic positive quaternary quadratic forms which represent all even integers, do not exceed a fixed upper bound B_2.*

Let $\Phi_1(x)$ in (2.01) represent all even integers. Let a_{11} be the minimum of Φ_1. Then

(2.31) $$a_{11} \leq 2.$$

Next

$$a_{11}\Phi_1 = \Phi_3 = z_1^2 + \phi_3(z_2, z_3, z_4)$$

where ϕ_3 is given by (2.02). Let a be the minimum of ϕ_3. Then, it is easily seen that

(2.32) $$a \leq \beta(a_{11})$$

where $\beta(a_{11})$ may be taken as the least multiple of $2a_{11}$, which is not a square. Proceeding further we see that

$$aa_{11}\Phi_1 = au_1^2 + u_2^2 + \Psi_4(u_3, u_4)$$

where Ψ_4 is given by (2.12). Let M be the minimum of Ψ_4. Then

(2.33) $$M \leq \Delta(a_{11}, a)$$

where $\Delta(a_{11}, a)$ may be taken as the least multiple of $2a_{11}a$ which is not represented by $au_1^2 + u_2^2$. Next

$$Maa_{11}\Phi_1 = Mau_1^2 + Mu_2^2 + U_3^2 + (ML - N^2)u_4^2$$

where

(2.341) $$ML - N^2 = aa_{11}^2 D, \qquad D = | a_{ij} |,$$

and

(2.342) $$D \leq ML - N^2 \leq B(M, a, a_{11}) \leqq B_2$$

where $B(M, a, a_{11})$ may be taken as a multiple of $2M \cdot a \cdot a_{11}$ which is not represented by

(2.34) $$Mau_1^2 + Mu_2^2 + U_3^2.$$

We are now ready to prove that B_2 in (2.342) is an absolute constant.

One may easily verify that $\beta(1) = 2$ and $\beta(2) = 8$. Examining the form $au_1^2 + u_2^2$ we get the following values for $\Delta(a_{11}, a)$ in (2.33):

a_{11}	1	1	2	2	2	2	2	2	2	2
a	1	2	1	2	3	4	5	6	7	8
$\Delta(a_{11}, a)$	6	20	12	40	24	48	40	48	84	160

To extract the best value of B_2 out of the above inequalities one should determine the best value of $B(M, a, a_{11})$ for each set of values M, a_1, a_{11} permitted by this table. Although this presents no difficulty, the computation is somewhat lengthy and we shall therefore merely prove the existence of such an upper bound. To do this it suffices, in view of the above discussion, to show that in every case the ternary form in (2.34) will fail to represent a multiple of $2Maa_{11}$. This last follows at once from a result due to Hasse.[3]

3. The canonical form (C_p). We shall find the following normalization useful in the subsequent discussion.

THEOREM 3. *Every properly primitive classic quaternary quadratic form with integral coefficients and invariants*[4] o_μ *is equivalent to a canonical form*

(3.01) $$f = \Sigma a_{ij}x_ix_j = x'Ax$$

of determinant $| A | = | a_{ij} |$ *whose leading principal minors are*

$$a_{11} = A_1, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = o_1 A_2, \qquad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = o_1^2 o_2 A_3$$

---

[3] Hasse (6a), § 11.

[4] After Minkowski and Smith. We employ the notation of Minkowski.

where

(C_p) $\qquad A_\mu$ *or* $\tfrac{1}{2}A_\mu$ *is an odd prime not dividing* $|a_{ij}|A_kA_l$

$$(\mu, k, l) = (1,2,3),\ (2,1,3),\ (3,1,2).$$

Since our form is properly primitive we may assume at once that a_{11} is an odd prime not dividing $|a_{ij}|$. Write $o_1^2o_2A_{ij}$ for the algebraic complement of a_{ij} in $|a_{ij}|$. Then $F = \Sigma A_{ij}X_iX_j$ is the reciprocal [5] of f. In view of the choice of a_{11} the ternary section $F(0, X_2, X_3, X_4)$ of F is a primitive ternary form of invariants [6]

$$\Omega = o_3 \quad \text{and} \quad \Delta = o_2a_{11}. \tag{3.03}$$

This ternary, however, is equivalent to a form whose third coefficient A_{44} $(=A_3)$ and the leading coefficient $(=A_2)$ of whose reciprocal are distinct odd primes not dividing $o_1o_2o_3A_1$ or doubles of such primes.[7] Replacement of the ternary section by this canonical form does not disturb [8] the choice of a_{11}.

4. The associates of a given quadratic form. In this section we assume that our quadratic form (2.01) is a canonical form (C_p). Multiplying through by $a_{11} = A_1$ we obtain

$$A_1f = X_1^2 + o_1\sum_{i,j=2}^{4}\alpha_{ij}^{(1)}x_ix_j, \tag{4.01}$$

where

$$o_1\alpha_{ij}^{(1)} = \begin{vmatrix} a_{11} & a_{1j} \\ a_{i1} & a_{ij} \end{vmatrix} \quad \alpha_{22}^{(1)} = A_2,\ X_1 = a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4. \tag{4.02}$$

Next

$$A_2A_1f = A_2X_1^2 + o_1(X_2^2 + a_{11}o_2\sum_{i,j=3}^{4}\alpha_{ij}^{(2)}x_ix_j), \tag{4.03}$$

where, in view of a determinant theorem of Sylvester,

$$\begin{vmatrix} \alpha_{22}^{(1)} & \alpha_{2j}^{(1)} \\ \alpha_{i2}^{(1)} & \alpha_{ij}^{(1)} \end{vmatrix} = \frac{1}{o_1^2}\begin{vmatrix} o_1\alpha_{22}^{(1)} & o_1\alpha_{2j}^{(1)} \\ o_1\alpha_{i2}^{(1)} & o_1\alpha_{ij}^{(1)} \end{vmatrix} \tag{4.04}$$

$$= \frac{a_{11}}{o_1^2}\begin{vmatrix} a_{11} & a_{12} & a_{1j} \\ a_{21} & a_{22} & a_{2j} \\ a_{i1} & a_{i2} & a_{ij} \end{vmatrix} = \frac{a_{11}o_1^2o_2\alpha_{ij}^{(2)}}{o_1^2} = a_{11}o_2\alpha_{ij}^{(2)}.$$

$$\alpha_{33}^{(2)} = A_3 \quad \text{and} \quad X_2 = \alpha_{22}^{(1)}x_2 + \alpha_{23}^{(1)}x_3 + \alpha_{24}^{(1)}x_4. \tag{4.05}$$

Finally

---

[5] Cf. Dickson (3).

[6] Minkowski (10), Ch. XVIII.

[7] Ross (16). The precise statement of the theorem referred to implies that if Ω, Δ are odd A_2 and A_3 may be taken as odd primes.

[8] Cf. Dickson (3).

(4.06) $A_3A_2A_1f = A_3A_2X_1^2 + o_1A_3X_2^2 + o_1o_2A_1X_3^2 + o_1o_2o_3A_1A_2X_4^2,$
since

(4.07) $$\begin{vmatrix} \alpha_{33}^{(2)} & \alpha_{34}^{(2)} \\ \alpha_{43}^{(2)} & \alpha_{44}^{(2)} \end{vmatrix} = \frac{1}{(o_1^2o_2)^2}\begin{vmatrix} o_1^2o_2\alpha_{33}^{(2)} & o_1^2o_2\alpha_{34}^{(2)} \\ o_1^2o_2\alpha_{43}^{(2)} & o_1^2o_2\alpha_{44}^{(2)} \end{vmatrix}$$
$$= \frac{o_1\alpha_{22}^{(1)}\,|\,A\,|}{o_1^4o_2^2} = \frac{A_2o_1^4o_2^2o_3}{o_1^4o_2^2} = o_3A_2.$$
Here

(4.08) $$X_3 = \alpha_{33}^{(2)}x_3 + \alpha_{34}^{(2)}x_4, \quad X_4 = x_4.$$

We now introduce the form

(4.09) $G(X_1, X_2, X_3, X_4) = A_3A_2X_1^2 + o_1A_3X_2^2 + o_1o_2A_1X_3^2 + o_1o_2o_3A_1A_2X_4^2$

in the independent variables X_1, X_2, X_3, X_4. We shall call G the *associate* of f.

4.1. The form G is of interest by virtue of the following:

THEOREM 4.1. *Let f be a properly primitive quaternary quadratic form. Employ the notation of Theorem 3 and assume that f is in the canonical form of type (C_2). Then if f represents an integer m its associate G in* (4.09) *represents $A_1A_2A_3m$. Conversely, if the form G represents $A_1A_2A_3m$ and* (4.14) *and* (4.16) *hold, then the original form f represents m.*

The first part of the theorem is trivial, for if x_1, x_2, x_3, x_4 are integers, then by (4.02), (4.05), and (4.08), so also are X_1, X_2, X_3, X_4 and $A_1A_2A_3m$ is represented by G in view of (4.06).

Now let $G(X_1, X_2, X_3, X_4) = A_1A_2A_3m$. We seek integers x_1, x_2, x_3, x_4 such that $f(x_1, x_2, x_3, x_4) = m$. We take $x_4 = X_4$. By (4.09),

(4.11) $A_1A_2A_3m = A_3A_2X_1^2 + o_1A_3X_2^2 + o_1o_2A_1X_3^2 + o_1o_2o_3A_1A_2x_4^2,$

and hence
$$o_1o_2A_1X_3^2 + o_1o_2o_3A_1A_2x_4 \equiv 0 \pmod{A_3}.$$
But, by (4.07),

(4.12) $$o_3A_2 \equiv -[\alpha_{34}^{(2)}]^2 \equiv -s^2 \pmod{A_3},$$
whence
$$o_1o_2A_1(X_3^2 - s^2x_4^2) \equiv 0 \pmod{A_3}.$$
If

(4.14) $$(o_1o_2A_1, A_3) = 1,$$
then
$$(X_3 - sx_4)(X_3 + sx_4) \equiv X_3^2 - s^2x_4^2 \equiv 0 \pmod{A_3},$$
since A_3 is either an odd prime or double such a prime, we have
$$X_3 \equiv sx_4 \quad \text{or} \quad -X_3 \equiv sx_4 \pmod{A_3},$$

whence, replacing s by its value in (4.12), we get

$$\pm X_3 = A_3 x_3 + \alpha_{34}{}^{(2)} x_4$$

with integral x_3. Substituting the resulting value of $X_3{}^2$ into (4.11) we get

$$(4.15)\quad A_3A_2A_1 m = A_3A_2X_1{}^2 + o_1A_3X_2{}^2 + o_1o_2A_1[(A_3x_3 + \alpha_{34}{}^{(2)}x_4)^2 + o_3A_2x_4{}^2].$$

Replacing o_3A_2 by its value in (4.07), squaring the expression in the parenthesis, combining the similar terms, and dividing both members of (4.15) by the factor A_3 common to all terms, we get

$$(4.151)\quad A_2A_1 m = A_2X_1{}^2 + o_1X_2{}^2 + o_1o_2A_1\Big(\sum_{i,j=3}^{4} \alpha_{ij}{}^{(2)} x_i x_j\Big).$$

This equality, in turn, implies that

$$o_1X_2{}^2 + o_1o_2A_1\Big(\sum_{i,j=3}^{4} \alpha_{ij}{}^{(2)} x_i x_j\Big) \equiv 0 \pmod{A_2}.$$

But, by (4.04),

$$o_2A_1\alpha_{ij}{}^{(2)} \equiv -\alpha_{2i}{}^{(1)}\alpha_{2j}{}^{(1)} \pmod{A_2},$$

and hence

$$\begin{aligned} o_2A_1 \sum_{i,j=3}^{4} \alpha_{ij}{}^{(2)} x_i x_j &= o_2A_1(\alpha_{33}{}^{(2)}x_3{}^2 + 2\alpha_{34}{}^{(2)}x_3x_4 + \alpha_{44}{}^{(2)}x_4{}^2) \\ &\equiv -[(\alpha_{23}{}^{(1)})^2x_3{}^2 + 2\alpha_{23}{}^{(1)}\alpha_{24}{}^{(1)}x_3x_4 + (\alpha_{24}{}^{(1)})^2x_4{}^2] \\ &\qquad\qquad\qquad\qquad\qquad\qquad \pmod{A_2} \\ &\equiv -(\alpha_{23}{}^{(1)}x_3 + \alpha_{34}{}^{(1)}x_4)^2 \pmod{A_2}. \end{aligned}$$

Thus, (4.151) becomes

$$o_1[X_2{}^2 - (\alpha_{23}{}^{(1)}x_3 + \alpha_{24}{}^{(1)}x_4)^2] \equiv 0 \pmod{A_2}.$$

If

$$(4.16)\quad (o_1, A_2) = 1,$$

then

$$\begin{aligned} (X_2 - \alpha_{23}{}^{(1)}x_3 - \alpha_{24}{}^{(1)}x_4)&(X_2 + \alpha_{23}{}^{(1)}x_3 + \alpha_{24}{}^{(1)}x_4) \\ &= X_2{}^2 - (\alpha_{23}{}^{(1)}x_3 + \alpha_{24}{}^{(1)}x_4)^2 \equiv 0 \pmod{A_2}, \end{aligned}$$

and since A_2 is a prime or a double of a prime, we have

$$X_2 = A_2x_2 + \alpha_{23}{}^{(1)}x_3 + \alpha_{24}{}^{(1)}x_4 \quad\text{or}\quad -X_2 = A_2x_2 + \alpha_{23}{}^{(1)}x_3 + \alpha_{24}{}^{(1)}x_4,$$

where x_2 is an integer. Substituting the resulting value of $X_2{}^2$ into (4.151), we get

$$(4.17)\quad \begin{aligned} A_2A_1 m = A_2X_1{}^2 &+ o_1(A_2x_2 + \alpha_{23}{}^{(1)}x_3 + \alpha_{24}{}^{(1)}x_4)^2 \\ &+ o_1o_2A_1\Big(\sum_{i,j=3}^{4} \alpha_{ij}{}^{(2)} x_i x_j\Big). \end{aligned}$$

Replacing $A_1o_2\alpha_{ij}{}^{(2)}$ by their values in (4.04), squaring the expression in the

parenthesis, combining similar terms, and dividing both members of (4.17) by the factor A_2 common to all terms, we get

$$(4.171) \qquad A_1 m = X_1^2 + o_1 \sum_{i,j=2}^{4} \alpha_{ij}^{(1)} x_i x_j.$$

Again, the last equality implies that

$$(4.172) \qquad X_1^2 + o_1 \sum_{i,j=2}^{4} \alpha_{ij}^{(1)} x_i x_j \equiv 0 \pmod{A_1}.$$

But, by (4.02), $o_1 \alpha_{ij}^{(1)} \equiv -a_{1i} a_{1j} \pmod{A_1}$ and hence

$$o_1 \sum_{i,j=2}^{4} \alpha_{ij}^{(1)} x_i x_j \equiv -\sum_{i,j=2}^{4} a_{1i} a_{1j} x_i x_j = -(a_{12}x_2 + a_{13}x_3 + a_{14}x_4)^2 \pmod{A_1}.$$

Thus, (4.172) becomes

$$(X_1 - a_{12}x_2 - a_{13}x_3 - a_{14}x_4)(X_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4) \equiv X_1^2 - (a_{12}x_2 + a_{13}x_3 + a_{14}x_4)^2 \equiv 0 \pmod{A_1}.$$

Since A_1 is a prime, we have

$$X_1 \text{ or } -X_1 = A_1 x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4$$

for an integer x_1. Substituting the resulting value of X_1^2 into (4.171), we get

$$(4.173) \quad A_1 m = (A_1 x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4)^2 + o_1 \sum_{i,j=2}^{4} \alpha_{ij}^{(1)} x_i x_j.$$

Replacing $o_1 \alpha_{ij}^{(1)}$ by their values in (4.02), squaring the expression in the parenthesis, combining similar terms and dividing both members of (4.173) by the factor A_1 common to all terms, we get

$$m = \sum_{i,j=1}^{4} a_{ij} x_i x_j$$

with integral $x_1, \cdots, x_4$. Thus m is represented by f.

FORMS OF ODD DETERMINANTS.

5. A set of sufficient conditions in terms of generic characters. We shall now restrict ourselves to the study of properly primitive forms (2.01) of odd determinants. We shall assume that such a form f is already in a canonical form of type (C_p). Then, since in this case A_1, o_1, o_2, o_3 are all odd, the conditions (4.14) and (4.16) of Theorem 4.1 hold in view of the choice of A_1, A_2, A_3. Thus if the associate G of f represents the multiple $A_1A_2A_3m$ of m then f represents m. Consequently should G represent all integers $\geq K$, and

therefore all multiples $A_1A_2A_3m \geqq K$ of $A_1A_2A_3$ then f would represent all integers $m \geqq K/A_1A_2A_3$. But the form G is of the type considered by Kloosterman (7). We may therefore apply to the form G Kloosterman's conditions 1°-5° assuring representation of all large integers.[9]

We note that A_1 is an odd prime. The same may be assumed in this case of A_2 and A_3 by the proof of Theorem 3.0. Our choice of A_1, A_2, A_3 implies that 4° and 5° hold. Condition 3° becomes

$$(5.1) \qquad o_1 = 1$$

and, if we write ω_2 for an odd prime factor of o_2, condition 2° becomes

(5.21) $(A_2|\omega_2) = (-1|\omega_2)$ or $(A_2|\omega_2) = (-o_3|\omega_2)$, or both, if $\omega_2{}^2 \nmid o_2$ and $\omega_2 \nmid$

(5.22) $(A_2|\omega_2) = (-1|\omega_2)$, if either (1) $\omega_2{}^2|o_2$ or (2) $\omega_2|o_3$;

(5.23) $(-o_3A_2|A_3) = 1, \ (-o_2A_1A_3|A_2) = 1, \ (-A_2|A_1) = 1.$

The condition (5.23) is satisfied by all forms, for, by virtue of (4.07), (4.02) and (4.04), we have

$$-o_3A_2 \equiv (\alpha_{34}{}^{(2)})^2 (\text{mod } A_3), \quad -A_2 \equiv a_{12}{}^2 \ (\text{mod } A_1),$$
$$-o_2A_1A_3 \equiv (\alpha_{23}{}^{(1)})^2 \ (\text{mod } A_2).$$

The condition (5.1) restricts the value of the first invariant o_1 of f. Next, in view of the choice of A_2 and the definition of the generic characters of f, it is clear that the relations (5.21) and (5.22) are in fact conditions upon the generic characters of f with respect to the odd prime factors of o_2.

The condition (5.21) may be modified by virtue of the following considerations. If $(o_3|\omega_2) = -1$, then $(-1|\omega_2)$ and $(-o_3|\omega_2)$ have opposite signs and $(A_2|\omega_2)$ must be equal to one or to the other, and hence at least one (and, of course, only one) of the relations in (5.21) holds true. If, however, $(o_3|\omega_2) = 1$, then $(-1|\omega_2) = (-o_3|\omega_2)$ and the two relations in (5.21) coincide and, therefore, either both hold true or both fail according as $(A_2|\omega_2) = (-1|\omega_2)$ or $(A_2|\omega_2) = -(-1|\omega_2)$. Thus (5.21) may be replaced by

(5.24) $\qquad (A_2|\omega_2) = (-1|\omega_2)$ if $\omega_2{}^2 \nmid o_2, \ \omega_2 \nmid o_3, \ (o_3|\omega_2) = 1.$

The part[10] $(\mu_a, \mu_b, \mu_c, \mu_d) = (0, 0, 0, 0)$ of the condition 1° is in fact the necessary and sufficient condition in order that a form (1.1) with odd coefficients a, b, c, d should fail to represent zero properly modulo 8. In view of the choice of A_1, A_2, A_3 and the formulae (4.02), (4.05), (4.08), the form f

---

[9] (7), Section 4.6, p. 453.
[10] Kloosterman (7), p. 453.

in (3.01) and its associate G in (4.09) either both represent zero properly modulo 8 or both fail to do so. The above mentioned conditions

(5.25) $\quad a \equiv b \equiv c \equiv d \pmod 4 \quad a + b + c + d \equiv 4 \pmod 8$

as applied to G, yield conditions

(5.26) $\quad o_1 \equiv o_3 \pmod 8,\ o_3 A_2 \equiv 1 \pmod 4,\ o_2 A_1 A_3 \equiv 1 \pmod 4.$

For, (5.251) becomes $A_3 A_2 \equiv o_1 A_3 \equiv o_1 o_2 A_1 \equiv o_1 o_2 o_3 A_1 A_2 \pmod 4$, and hence $A_2 \equiv o_1$, $A_3 \equiv o_2 A_1$, $1 \equiv o_3 A_2 \pmod 4$. The first and the third of these last congruences imply $o_1 \equiv o_3 \pmod 4$. Moreover, (5.252) becomes

$$A_3(A_2 + o_1) + o_1 o_2 A_1 (1 + o_3 A_2) \equiv 4 \pmod 8.$$

Since $A_2 + o_1 \equiv 2 o_1 \equiv 2$ and $1 + o_3 A_2 \equiv 2 \pmod 4$, each of the two terms above is double an odd integer and the last congruence together with $A_3 \equiv o_2 A_1 \pmod 4$ implies $A_3(A_2 + o_1) \equiv o_1 A_3 (1 + o_3 A_2) \pmod 8$. Therefore $A_2 + o_1 \equiv o_1 + o_1 o_3 A_2$ and $1 \equiv o_1 o_3 \pmod 8$. It is easily seen that conditions (5.26) imply that G satisfies (5.25).

The conditions (5.262) and (5.263) are equivalent to

(5.27) $$\Psi = (-1)^{\frac{1}{2}(o_3 A_2 + 1) \cdot \frac{1}{2}(o_2 A_1 A_3 + 1)} = -1.$$

Since the ternary form $F(0, X_2, X_3 X_4)$ in Section **3**, has invariants $\Omega = o_2$ and $\Delta = o_2 A_1$, and since A_3 and A_2 are represented simultaneously by this ternary and its reciprocal we have[11]

$$\Psi \cdot (A_3 | o'_3)(A_2 | o'_2 A_1) = (-1)^{\frac{1}{2}(o_3+1) \cdot \frac{1}{2}(o_2 A_1 + 1)}.$$

Here $o_i = o_i' o_i''^2$, and $o_i''^2$ is the largest square dividing o_i. In view of (4.02),

$$(A_2 | A_1) = (-o_1 | A_1) = (-1)^{[(A_1-1)/2] \cdot [(o_1+1)/2]} (A_1 | o_1'),$$

and therefore

$$(A_2 | o'_2 A_1) = (A_2 | o'_2)(A_1 | o'_1)(-1)^{\frac{1}{2}(A_1-1)\frac{1}{2}(o_1+1)}.$$

Since (5.261) implies that $\frac{1}{2}(o_3 + 1)\frac{1}{2}(o_3 A_1 + 1) + \frac{1}{2}(o_1 + 1)\frac{1}{2}(A_1 - 1) \equiv \frac{1}{2}(o_3 + 1)$ modulo 2, we have

(5.28) $$(A_3 | o'_3)(A_2 | o'_2)(A_1 | o'_1) = \Psi \cdot (-1)^{\frac{1}{2}(o_3+1)} = -(-1)^{\frac{1}{2}(o_3+1)}$$

in view of (5.27).

We see that conditions (5.26) are equivalent to (5.261) and (5.28). In view of the choice of A_1, A_2, A_3 condition (5.28) is a restriction upon the generic characters of f.

---

[11] Smith, vol. 1, p. 470.

If (5.1) holds, then, in view of (5.261), (5.28) becomes

$$(A_3 | o'_3)(A_2 | o'_2) = 1. \tag{5.29}$$

Employing the notation of Minkowski,[10] we may state our conclusions in the following form:

THEOREM 5. *Let ϕ be a properly primitive quaternary form of odd determinant and the invariants o_1, o_2, o_3. Let ω_2 be an odd prime factor of o_2. Let finally*

$$o_1 = 1, \tag{5.1}$$

(5.31) $(\phi_2 | \omega_2) = (-1 | \omega_2)$ *for* ω_2 *such that* $\omega_2^2 \nmid o_2 o_3$, $(o_3 | \omega_2) = 1$,

(5.32) $(\phi_2 | \omega_2) = (-1 | \omega_2)$ *for* ω_2 *such that* $\omega_2^2 | o_2 o_3$,

(5.33) $(\phi_3 | o'_3)(\phi_2 | o'_2) = -1$, *if* $o_3 \equiv 1 \pmod 8$.

Then the form ϕ represents all but a finite number of integers.

The truth of this theorem follows at once from the fact that conditions (5.1), (5.31)-(5.33) imply that (5.1), (5.22)-(5.24) but not (5.26) hold for the associate G of a canonical form f of type (C_p) in the class of ϕ and hence G represents all but a finite number of integers.

6. The necessity of conditions (5.1) and (5.32). We ask now if the conditions (5.1), (5.31), (5.32), (5.33) are necessary in order that ϕ should represent all integers with but a finite number of exceptions. We find at once that (5.1) and (5.32) are necessary. For, should there be an odd prime divisor $\omega_1 > 1$ of o_1, ϕ would not represent integers m such that

$$(m, \omega_1) = 1, \qquad (m | \omega_1) = -(\phi_1 | \omega_1). \tag{6.1}$$

(in view of (4.11)). Next, suppose that (5.32) should fail. Then $(\phi_2 | \omega_2) = -(-1 | \omega_2)$. First let $\omega_2^2 | o_2$. Then, in view of (4.11), with $o_1 = 1$, and the choice of A_2, A_1, ϕ would not represent integers m such that

$$m = \omega_2 m_1, \qquad (m_1, \omega_2) = 1. \tag{6.2}$$

Next let $\omega_2^2 \nmid o_2$, but $\omega_2 | o_3$. Then by (4.11), with $o_1 = 1$, ϕ would not represent integers m such that

$$m = \omega_2 m_1, \quad (m_1, \omega_2) = 1, \quad (m_1 | \omega_2) = -(\phi_3 | \omega_2)(\phi_2 | \omega_2)(\bar{o}_2 | \omega_2), \tag{6.3}$$

where $o_2 = \omega_2 \bar{o}_2$.

One sees without any difficulty that there are infinitely many integers of any one of the types (6.1), (6.2), and (6.3).

7. The conditions (5.31) and (5.33). The condition (5.31) differs essentially from (5.1) and (5.32). Its failure does not directly imply that there is an infinity of integers not represented, but rather that *if there is one integer not represented, then there is an infinity of such integers.* More precisely: let ϕ, and therefore f, represent an integer $\omega_2{}^2m$ where ω_2 is an odd prime factor of o_2 such that

$$(7.1) \qquad \omega_2{}^2 \nmid o_2 o_3, \quad (o_3 | \omega_2) = 1 \quad \text{and} \quad (\phi_2 | \omega_2) = -(-1 | \omega_2).$$

Then the condition (5.31), and hence (5.24), fails and we have

$$(7.11) \qquad (A_2 | \omega_2) = -(-1 | \omega_2) \quad \text{and} \quad (A_2 | \omega_2) = -(-o_3 | \omega_2).$$

Since $A_1 A_2 A_3 \omega_2{}^2 m$ is represented by G we have, in view of (4.11) with $o_1 = 1$,

$$A_3 A_2 X_1{}^2 + A_3 X^2 \equiv 0 \pmod{\omega_2},$$

and, since $(A_3, w_2) = 1$,

$$A_2 X_1{}^2 + X_2{}^2 \equiv 0 \pmod{\omega_2}.$$

Therefore, by (7.11$_1$) $X_1 \equiv X_2 \equiv 0 \pmod{\omega_2}$. Replacing X_1 and X_2 by $\omega_2 X'_1$ and $\omega_2 X'_2$ respectively, in (4.11), and dividing every term of both members of this equality by ω_2, we see that

$$\bar{o}_2 A_1 X_3{}^2 + \bar{o}_2 o_3 A_1 A_2 X_4{}^2 \equiv 0 \pmod{\omega_2},$$

and, since $(\bar{o}_2 A_1, \omega_2) = 1$,

$$X_3{}^2 + o_3 A_2 X_4{}^2 \equiv 0 \pmod{\omega_2}.$$

But by (7.11$_2$), $(-o_3 A_2 | \omega_2) = -1$, and hence $X_3 \equiv X_4 \equiv 0 \pmod{\omega_2}$. Write $X_3 = \omega_2 X'_3$, $X_4 = \omega_2 X'_4$. Then, dividing once more every term of both members of (4.11) by ω_2, we get

$$A_1 A_2 A_3 m = G(X'_1, X'_2, X'_3, X'_4).$$

Thus G represents $A_1 A_2 A_3 m$, and hence f, and therefore ϕ represents m. It follows, therefore, that if a form ϕ should not fulfil condition (5.31) then if it does not represent an integer m it also does not represent $\omega_2{}^2 m$ and in general $\omega_2{}^{2\kappa} m$ for every integer κ. This proves the above statement in italics. It is seen then that *such a form ϕ either represents all integers or there are infinitely many integers not represented by ϕ.* We shall speak of the set of integers of the form $\omega_2{}^{2\kappa} m$ as the *tower* $\omega_2{}^{2\kappa} m$ or the tower generated by ω_2 and m.

In view of the discussion preceding the statement of Theorem 5, the failure of condition (5.33) implies that if f does not represent an *even* integer

$2m$, then it does not represent the whole tower $2^{2\kappa+1}m$, that is, it does not represent an infinity of integers.

8. Forms representing all large integers and not covered by Theorem 5. Such forms may be divided into two types. Type P_1, for which (5. 31) does not hold, and type P_3 for which (5. 31) holds true but (5. 33) fails to hold.

Forms of type P_3 must represent all even integers. For, if such a form fails to represent an even integer $2m$ then it fails to represent the whole tower $2^{2\kappa+1}m$. In view of Theorem 2. 3, the determinant of such a form does not exceed B_2. Thus, there is only a finite number of classes of forms of type P_3.

Forms of type P_1 must represent all integers. For, if such a form fails to represent an integer m then it fails to represent the whole tower $\omega^{2\kappa}m$. Therefore, by Theorem 2. 0, the determinant of such a form does not exceed 112.

We shall determine all forms of type P_1. The only odd determinants $D < 112$ which permit such forms are given together with the desired invariants o_2, o_3 ($(o_3 | \omega_2) = 1$) by the following table.

(8. 01)

D	9	63	25	49
$\omega_2 = o_2$	3	3	5	7
o_3	1	7	1	1

Since every form ϕ under consideration which does not represent 1 also does not represent $\omega_2{}^{2\kappa}$, we need consider only forms which represent 1. Such forms are equivalent to

$$(8.02)\quad \phi = \xi^2 + ax^2 + by^2 + cz^2 + 2ryz + 2sxz + 2txy = \xi^2 + \mu(x, y, z).$$

If ϕ is a form of one of the determinants given in the table (8. 01) and if it possesses the indicated invariants o_2 and o_3 then $\mu(x, y, z)$ is a properly primitive form with invariants $\Omega = o_2$ and $\Delta = o_3$ and odd determinant $\Omega^2\Delta = o_2{}^2 o_3 = D$. Since (7. 1) holds, we have

$$(8.03)\qquad (\mu | \omega_2) = -(-1 | \omega_2)$$

for the generic character $(\mu | \omega_2)$ of μ. For, if we choose a to be any integer prime to ω_2 and properly represented by μ then, in view of (8. 02),

$$(\phi_2 | \omega_2) = (a | \omega_2) = (\mu | \omega_2).$$

We may assume next that μ in (8. 02) is a reduced form and we need to consider all forms (8. 02) in which μ is a reduced form of the above mentioned invariants and genus satisfying (8. 03). We list such forms in the following table.

D	$\Omega=\omega_2$	Δ		a	b	c	r	s	t	$(\mu\|\omega_2)$	$(-1\|\omega_2)$
9	3	1	$\mu^{(1)}$	1	3	3	0	0	0	1	-1
25	5	1	$\mu^{(2)}$	2	3	5	0	0	-1	-1	1
49	7	1	$\mu^{(3)}$	1	7	7	0	0	0	1	-1
			$\mu^{(4)}$	2	4	7	0	0	-1		
63	3	7	$\mu^{(5)}$	1	3	21	0	0	0	1	-1
			$\mu^{(6)}$	1	6	12	-3	0	0		
			$\mu^{(7)}$	3	3	7	0	0	0		

We write

$$\phi^{(i)}=\xi^2+\mu^{(i)} \qquad (i=1,\cdots,7).$$

Each of the forms $\phi^{(1)},\cdots,\phi^{(7)}$ satisfies the conditions (5.1), (5.32), but not the condition (5.31). Hence each of these forms either represents all integers or fails to represent an infinite number of integers (Cf. 7). One sees at once that

(8.1) $$\phi^{(3)}\neq 3,\quad \phi^{(5)}\neq 6,\quad \phi^{(6)}\neq 3,\quad \phi^{(7)}\neq 2,$$

and therefore these forms belong to the second category, i. e., they do not represent an infinity of integers. The form $\phi^{(1)}=\xi^2+x^2+3y^2+3z^2$ as is well known,[12] represents all integers. Consider next the form

$$\phi^{(2)}(\xi,x,y,z)=\xi^2+2x^2+3y^2+5z^2-2xy.$$

Its ternary section $\psi=\phi(\xi,x,y,0)=\xi^2+2x^2+3y^2-2xy$ belongs to a genus of one class[13] of determinant 5. Its invariants are $\Omega=1$, $\Delta=5$, and its character is

$$(\Psi|5)=(2|5)=-1=-(-\Omega|5)$$

where Ψ is the reciprocal of ψ. Therefore[14] integers not represented by ψ are

(8.11) $$5^{2\kappa+1}(5n+1),\quad 5^{2\kappa+1}(5n+4).$$

In order to prove that $\phi^{(3)}$ represents all integers we need only prove that it represents all integers of the form (8.11). But these integers are represented by the ternary section $\phi^{(2)}(\xi,x,0,z)=\xi^2+2x^2+5z^2$ which represents all integers not of the form[15] $5^{2k+1}(5n+2)$, $5^{2k+1}(5n+3)$ and hence all integers of the form (8.11). Thus $\phi^{(2)}$ represents all integers.

Finally, consider the form

$$\phi^{(4)}=\xi^2+2x^2+4y^2+7z^2-2xy.$$

---

[12] Ramanujan (14), Dickson (4), pp. 111-113.
[13] Jones (6), Borissow (1).
[14] Ross (17), Lemmas 1-3.
[15] Ramanujan (14), Dickson (4), pp. 111-113.

Since $2\phi^{(4)} = 2\xi^2 + (2x - y)^2 + 7y^2 + 14z^2$, the form $\phi^{(4)}$ will represent all integers if

$$\psi^{(4)} = X^2 + 2\xi^2 + 7y^2 + 14z^2$$

will represent all even integers. The ternary section

$$X^2 + 2\xi^2 + 7y^2 \tag{8.12}$$

represents [16] all even integers $\equiv 0$ or $1 \pmod 3$ which are not of the form

$$7^{2\kappa+1}(14m + R); \qquad R = 10,\ 12, \text{ and } 6. \tag{8.13}$$

If an integer is $\equiv 0$ or $1 \pmod 3$ and is of the form (8.13), we need concern ourselves only with those of type $L = 7(14m + R)$. Since at least one of the integers

$$L - 14 \cdot 3^2 = 7[14(m-1) + (R-4)]$$

or

$$L - 14 \cdot 6^2 = 7[14(m-5) + (R-2)]$$

is not of the form (8.13) and neither one is $\equiv 2 \pmod 3$, one of them is represented by (8.12), and hence L, and therefore $7^{2\kappa}L$, is represented by $\phi^{(4)}$.

Finally, let an integer be of the form $3n + 2$. If $z \not\equiv 0 \pmod 3$ then $3n + 2 - 14z^2 \equiv 3n + 2 - 2 \equiv 0 \pmod 3$. We may again assume that $7^2 \nmid 3n + 2$. If $(7, 3n + 2) = 1$, then $3n + 2 - 14$ is not of the form (8.13) and hence is represented by (8.12). If $7 \mid 3n + 2$, then

$$3n + 2 = 7(14m + M), \qquad M \not\equiv 0 \pmod 7.$$

In this case at least one of the integers

$$7(14m + M) - 14 = 7[14m + (M-2)],$$
$$7(14m + M) - 14 \cdot 2^2 = 7[14m + (M-8)],$$

or

$$7(14m + M) - 14 \cdot 4^2 = 7[14(m-2) + (M-4)]$$

is not of the form (8.13). The only numbers for which the above differences are negative, are ≤ 560 and are easily seen to be represented by $\phi^{(4)}$. Thus $\phi^{(4)}$ represents all integers.

We may now supplement Theorem 5 by

THEOREM 8. *Except for forms given in Theorem 5, there is only a finite number of classes of forms of odd determinants representing all large integers. These consist of a finite number of classes of forms of type P_3 and exactly*

---

[16] Pall (12).

three classes of type P_1, viz., the three classes containing $\phi^{(1)}$, $\phi^{(2)}$ and $\phi^{(4)}$. The classes to which $\phi^{(1)}$ and $\phi^{(2)}$ belong may be completely described by their generic invariants, for they are the only classes in their respective genera.[17] *This is not true of the class of $\phi^{(4)}$.*

9. One observes that $\phi^{(3)} = \xi^2 + \mu^{(3)}$ and $\phi^{(4)} = \xi^2 + \mu^{(4)}$ (Cf. the table in **8**) do not represent the same large integers even though they belong to the same genus. For, $\phi^{(4)}$ represents all large integers, whereas $\phi^{(3)}$ does not represent integers in the tower $3 \cdot 7^{2\kappa}$. Similarly the forms $\phi^{(5)}$, $\phi^{(6)}$, $\phi^{(7)}$ do not represent the same large integers in view of (8.1). It appears that when a quaternary form f fails to represent zero properly modulo p^{α} where $\alpha \geq 2$ and p is a prime, then the behavior of f for large integers depends not only upon the values of its generic invariants but also upon its accidental behavior for small values of the variables. For, should f fail to represent a small integer $p^{\nu}m$, it would not represent the whole tower $mp^{2\kappa+\nu}$. In our case when $p = \omega_2$ then $\nu = 0$ and when $p = 2$ then $\nu = 1$.

The failure of (5.31) or (5.33) implied that our form was not a zero form modulo $\omega_2{}^2$ or modulo 8 respectively.

ST. LOUIS UNIVERSITY.

BIBLIOGRAPHY.

1. E. Borissow, *Reduction of Positive Ternary Quadratic Forms by Selling's Method, with a Table of Reduced Forms for all Determinants from 1 to 200*, St. Petersburg (1890), pp. 1-108; tables 1-116. (Russian).

2. L. E. Dickson, *American Journal of Mathematics*, vol. 49 (1927), pp. 39-56.

3. ———, *Studies in the Theory of Numbers*, University of Chicago Press, 1930.

4. ———, *Modern Elementary Theory of Numbers*, University of Chicago Press, 1939.

5. Jacobi, *Werke*, vol. 6, pp. 281-302.

6. B. W. Jones, *A Table of Eisenstein—reduced Positive Ternary Quadratic Forms of determinant* $\leqq 200$, (1935), Bulletin No. 97 of the National Research Council.

6a. H. Hasse, "Über die Darstellbarkeit von Zahlen durch quadratische Formen im Körper der rationalen Zahlen," *Journal für Mathematik*, vol. 152 (1923), pp. 205-224.

7. H. D. Kloosterman, *Acta Mathematica*, vol. 49 (1926), pp. 454-464.

8. J. L. Lagrange, *Oeuvres*, vol. 3 (1869), pp. 189-201.

9. J. Liouville. See, for example, *Journal de Mathématiques* (2), vol. 1 (1856), p. 230.

---

[17] Townes (20).

10. H. Minkowski, "Grundlagen für eine Theorie der quadratischen Formen," *Werke*, Bd. 1, pp. 3-144.

11. D. C. Morrow, *University of Chicago Thesis*, 1928.

12. G. Pall, *Bulletin of the American Mathematical Society*, vol. 46 (1940), p. 291.

13. T. Pepin, *Atti Accad. Pont. Nuovi Lincei*, vol. 38 (1884-5), pp. 171-196, and elsewhere.

14. S. Ramanujan, *Proceedings of the Cambridge Philosophical Society*, vol. 19, I. (1916), pp. 11-21. Also *Collected Papers*, pp. 169-178. Cf. also Dickson (4), pp. 104-5. It was first pointed out by Dickson that one of the 55 forms originally listed by Ramanujan was not universal.

15. A. E. Ross, *Proceedings of the National Academy of Sciences*, vol. 18 (1932), pp. 600-608.

16. ———, *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 899-906.

17. ———, *American Journal of Mathematics*, vol. 55 (1933), pp. 293-302.

18. H. J. S. Smith, *Collected Papers*.

19. W. A. Tartakowsky, *Bull. Ak. Sc. U. R. S. S.*, vol. 7 (1929), pp. 111-122, 165-196.

20. S. B. Townes, *Annals of Mathematics*, vol. 41 (1940), pp. 57-58.

21. H. W. Turnbull, *The Theory of Determinants, Matrices, and Invariants*.

THE COMPLETION OF A PROBLEM OF KLOOSTERMAN.*

By Gordon Pall.

1. Introduction. The Euler-Lagrange proof of the theorem that every positive integer is a sum of four squares employed the fact that the form $x^2 + y^2 + z^2 + t^2$ is multiplicative. Hence Liouville [1] [1] examined the multiplicative forms $x^2 + ay^2 + bz^2 + abt^2$, and found the positive integers a, b, for which these forms represent all positive integers. Ramanujan [2] examined in similar fashion the forms

$$(1) \qquad f = (a, b, c, d) = ax^2 + by^2 + cz^2 + dt^2,$$

where a, b, c, d are positive integers, and $a \leqq b \leqq c \leqq d$. He found that there are 54 such forms which represent all positive integers; actually he had 55, and Dickson later pointed out his error [3]. More recently, Halmos found the 88 such forms which represent all positive integers with one exception [4].

Ramanujan, in the spirit of the analytic number theory which was then becoming popular, proposed and partly solved the more interesting problem of determining all forms (1) which represent all but a finite number of positive integers. Kloosterman [5] later solved this problem of Ramanujan, save that he was unable to decide whether the four forms

$$(2) \qquad (1, 2, 11, 38), \quad (1, 2, 17, 34), \quad (1, 2, 19, 22), \quad (1, 2, 19, 38)$$

represent all positive integers. In **11-13** we shall complete the problem by showing that these four forms do in fact represent all large integers. The technique used for this purpose is based on the fact that quadratic forms in the same genus have rational transformations into one another, which can be employed in particular cases to investigate the numbers represented integrally by the individual forms. Earlier similar attempts by the writer (6) failed because he did not then realize that Kloosterman's asymptotic formula could be adapted to settle the problem for numbers involving a limited power of 2 so that it is only necessary to consider, say, multiples of 4.

However, the most interesting feature of this paper consists in the elegant formulation of the conditions for a form to represent all large integers. This

* Received July 17, 1945.

[1] Numbers in square brackets refer to the references at the end of the paper.

formulation we owe largely to reading a manuscript of Arnold E. Ross, in which he considers the extension of Kloosterman's results to non-diagonal forms. Also, whereas much earlier work on quadratic forms is complicated by many cases involving the numerous invariants of the forms, we obtain our results here directly and simply by appealing to the arithmetical properties of the forms themselves. Although our Theorems 1, 2, and 4 are true, precisely as stated, for non-diagonal forms as well, this will not be proved here.

2. The pertinent properties of *f*. If f is to represent all large integers, then evidently the congruence $f \equiv n \pmod{k}$ must be solvable in integers x, y, z, t, for every pair of integers n and k, $k \neq 0$. If p is any prime we shall say that *f is p-adically universal*, when

$$ax^2 + by^2 + cz^2 + dt^2 \equiv n \pmod{p^r}, \tag{3}$$

is solvable in integers x, y, z, and t, for every n and r, $r \geqq 0$. Hence a necessary condition for f to represent all large integers is that f be p-adically universal for every p.

We shall later (Lemma 1) give precise criteria for such universality, as also (Lemma 2) for p-adic representation of zero.

We shall say that *f fails to represent zero p-adically*, if

$$ax^2 + by^2 + cz^2 + dt^2 \equiv 0 \pmod{p^r} \tag{4}$$

implies, for some r, that $x \equiv y \equiv z \equiv t \equiv 0 \pmod{p}$. If, however, (4) is solvable for every positive integer r in integers x, y, z, t not all divisible by p, we say that *f represents zero p-adically*. The connection with our problem is this. If f fails to represent zero p-adically for some p, suppose for a certain s, that $f \equiv 0 \pmod{p^s}$ implies $x \equiv y \equiv z \equiv t \equiv 0 \pmod{p}$. Then the number of representations of $p^{s+2k}n$ is the same as that of $p^s n$ in f, for every positive integer k. Hence *if* f does not represent one integer of the form $p^s n$, f does not represent infinitely many integers. And in any case the number of representations of the *large* number p^{s+2k} is bounded, so that, in the asymptotic formula soon to be encountered, $\chi(p, p^{s+2k})$ is near zero when k is large.

3. The principal theorems. Kloosterman's results are expressed in rather complicated fashion in terms of conditions on the coefficients of f. Interpreting his results by means of the notions of **2**, and using Lemmas 1 and 2, we get the following two theorems.

THEOREM 1. *If* (a) *f is p-adically universal for every p, and* (b) *f represents zero p-adically for every p, then f represents all sufficiently large integers.*

It will be observed that (a) is a necessary, and (b) is not a necessary condition.

THEOREM 2. *Of the forms f which are p-adically universal for every p, there are only a finite number of forms which fail to represent zero p-adically for some prime p_1, and yet represent all large integers.*

We shall in fact prove

THEOREM 3. *There are precisely 199 forms which fail to represent zero p-adically for some p_1, and yet represent all large integers. They are as follows, the first two having $p_1 = 3$ and 5, the others $p_1 = 2$:*

(A) $(1, 1, 3, 3)$;

(B) $(1, 2, 5, 10)$;

(C) $(1, 1, 5, 5)$, *and* $(1, 1, 1, t)$ $(t = 1, 9, 17, 25)$;

(D) $(1, 4, 5, 5)$, $(1, 1, 5, 20)$, $(1, 1, 4, t)$, $(1, 1, 1, 4t)$;

(E) $(1, 1, 10, 10)$, $(2, 2, 5, 5)$, $(1, r, 2, 2t)$, $(1, 17, 2, 2s)$, *and* $(1, 25, 2, 2r)$, $(r = 1, 9;\ s = 1, 9, 17)$;

(F) $(1, 1, 10, 40)$, $(5, 5, 2, 8)$, $(1, r, 2, 8t)$, $(1, r, 8, 2t)$, $(1, 17, 2, 8s)$, $(1, 17, 8, 2s)$, $(1, 25, 2, 8r)$, $(1, 25, 8, 2r)$;

(G) $(1, 4, 10, 10)$, $(2, 2, 5, 20)$, $(1, 4r, 2, 2t)$, $(4, r, 2, 2t)$, $(1, 68, 2, 2s)$, $(4, 17, 2, 2s)$, $(1, 100, 2, 2r)$, $(4, 25, 2, 2r)$;

(H) $(1, 4, 10, 40)$, $(2, 8, 5, 20)$, $(1, 4r, 2, 8t)$, $(1, 4r, 8, 2t)$, $(4, r, 2, 8t)$, $(4, r, 8, 2t)$, $(1, 68, 2, 8s)$, $(1, 68, 8, 2s)$, $(4, 17, 2, 8s)$, $(4, 17, 8, 2s)$, $(1, 100, 2, 8r)$, $(1, 100, 8, 2r)$, $(4, 25, 2, 8r)$, $(4, 25, 8, 2r)$;

(I) $(1, u, 2, 2v)$, $u, v = 3, 11,$ *and* 19;

(J) $(1, 4u, 2, 2v)$, $(4, u, 2, 2v)$;

(K) $(1, u, 2, 8v)$, $(1, u, 8, 2v)$;

(L) $(1, 4u, 2, 8v)$, $(1, 4u, 8, 2v)$, $(4, u, 2, 8v)$, $(4, u, 8, 2v)$.

Furthermore, we shall give in **4** a surprisingly easy proof of a theorem of which Theorem 1 is obviously a corollary:

THEOREM 4. *Let n denote any integer such that $f \equiv n \pmod{k}$ is solvable for every modulus k. For each prime p such that f fails to represent zero p-adically, impose an upper bound to the power of p in n. Then f represents every sufficiently large n (> 0) thus restricted.*

3a. Modification of the asymptotic formula. Kloosterman's formula [7] for the number of representations $f(n)$ of n by f is

$$(5)\qquad f(n)=\frac{\pi^2}{(abcd)^{1/2}}nS(n)+O(n^{17/18+\epsilon}).$$

To prove that $f(n)>0$ for n large, it suffices to show that

$$(6)\qquad S(n)>K/\log\log n,$$

where K is a positive constant depending only on f and the positive number ϵ. Kloosterman expresses $S(n)$ as a product over all primes p, namely

$$(7)\qquad S(n)=\prod_p \chi(p),$$

$$(8)\qquad \chi(p)=\chi(p,n)=1+A(p)+A(p^2)+\cdots,$$

$$(9)\qquad p^{4r}A(p^r)=\sum_h\sum_{x,y,z,t}\exp[2\pi ih(ax^2+by^2+cz^2+dt^2-n)/p^r],$$

where x, y, z, t range over all residues mod p^r, and h over all such residues prime to p. To put $\chi(p)$ into a more significant form, note that if $r\geqq 1$, the right member of (9) is

$$\begin{aligned}&\sum_{h,x,y,z,t \bmod p^r}\exp[2\pi ih(ax^2+by^2+cz^2+dt^2-n)/p^r]\\&\qquad -p^4\sum_{h_1,x,y,z,t \bmod p^{r-1}}\exp[2\pi ih_1(ax^2+by^2+cz^2+dt^2-n)/p^{r-1}]\\&\quad =p^rf(n,p^r)-p^{r+3}f(n,p^{r-1}),\end{aligned}$$

where $f(n,k)$ denotes the number of solutions of

$$(10)\qquad f(x,y,z,t)\equiv n \pmod{k}.$$

Hence

$$(11)\qquad \chi(p)=\lim_{r\to\infty}p^{-3r}f(n,p^r),$$

where it should be observed that if p^δ is the precise power of p in the determinant of f, $p^{-3r}f(n,p^r)$ is independent of r if $r\geqq\delta+3$ [8]. This interesting form of expression for $\chi(p)$ will be found in a paper by Tartakowsky [9], and in the work of Siegel.

We note that it is easy to prove by (11) that

$$(12)\qquad \text{if } p\nmid 2abcd,\qquad \chi(p)=(1-\epsilon_1p^{-2})\sum_{j=0}^{\nu}\epsilon_1{}^jp^{-j},$$

where $\epsilon_1=(abcd\,|\,p)$, $n=p^\nu n_1$, n_1 prime to p.

A sketch of Kloosterman's application of (5) is now in order. He proves easily that the product of the $\chi(p)$ over all primes save the finite number dividing $2abcd$, exceeds $K/\log\log n$ [10]. He then narrows the problem to forms whose coefficients satisfy certain conditions. Comparison with our Lemma 1 will show that these are the conditions for f to be universal for every p. Next, confining himself to such forms, he shows that for primes of a special kind, which by our Lemma 2 we now recognize as those for which f fails to represent zero p-adically, $\chi(p)$ is so near zero when n is divisible by a high power of p that (6) will not hold. This is to be expected from the last observation of **2**.

4. Proof of Theorem 4. We shall formulate the proof so as to apply to any m-ary quadratic form f for which a formula like (5) holds; $\chi(p)$ is then $\lim p^{-(m-1)r}f(n, p^r)$. Except possibly when $p = 2$ (see below) any f can be expressed modulo p^r as a sum of terms $p^{\alpha_i}a_ix_i^2$, say

$$(13) \qquad f \equiv p^{\alpha_1}a_1x_1^2 + \cdots + p^{\alpha_m}a_mx_m^2,$$
$$0 \leqq \alpha_1 \leqq \alpha_2 \leqq \cdots \leqq \alpha_m, \qquad a_1 \cdots a_m \text{ prime to } p.$$

To prove the theorem it is evidently sufficient to show that $\chi(p)$ is bounded away from zero for all large n, for each prime p dividing the determinant of f, and for the prime $p = 2$.

Case 1. Suppose n to be such that $f \equiv n \pmod{p^{\alpha_m+3}}$ is solvable with some x_i prime to p. We shall prove that $\chi(p) \geqq p^{-(m-1)(\alpha_m+3)}$. For, proceeding by induction, suppose $r \geqq \alpha_m + 3$ and that $f \equiv n \pmod{p^r}$ is solvable with, say, x_3 prime to p. The residues $x_1, x_2, x_4, \cdots, x_m \pmod{p^r}$ expand into p^{m-1} sets of residues mod p^{r+1}. For each of these we can choose h so that if x_3 is replaced by $x_3 + p^{r-\alpha_3}h$ if $p > 2$, or by $x_3 + 2^{r-\alpha_3-1}h$ if $p = 2$, then $f \equiv n \pmod{p^{r+1}}$ [11]. By repetitions of this process we see that $f \equiv n \pmod{p^r}$ has at least $p^{(m-1)(r-\alpha_m-3)}$ solutions, if $r \geqq \alpha_m + 3$. Hence $\chi(p) \geqq p^{-(m-1)(\alpha_m+3)}$.

Case 2. Let the power p^ν of p in n be bounded. Then if $f \equiv n \pmod{p^r}$ has p^ρ dividing every x_i but not $p^{\rho+1}$, then $2\rho \leqq \nu$ and the number of solutions of $f \equiv n \pmod{p^r}$ is $p^{m\rho}$ times the number of solutions of $f \equiv n/p^{2\rho} \pmod{p^{r-2\rho}}$. Hence by Case 1,

$$f(n, p^r) \geqq p^{m\rho}p^{(m-1)(r-2\rho-\alpha_m-3)}; \qquad \chi(p) \geqq p^{-\frac{1}{2}(m-2)\nu-(m-1)(\alpha_m+3)}.$$

Case 3. Finally, let f represent zero p-adically. The fact that $\chi(p)$ is

bounded away from zero follows from Case 2 if $p^{\alpha_m+3} \nmid n$. Let then $p^{\alpha_m+3} \mid n$. Now, $f \equiv 0 \pmod{p^{\alpha_m+3}}$ is solvable with an x_i prime to p. Hence $f \equiv n \pmod{p^{\alpha_m+3}}$ is similarly solvable, and Case 1 applies.

Although not needed in this article the remaining case with $p = 2$ will be resolved. We then have $f \equiv 2^\alpha(jx_1^2 + x_1x_2 + jx_2^2) +$ terms in other variables $\pmod{2^r}$, and $j = 0$ or 1. Hence, if $f \equiv n \pmod{2^r}$ is solvable with (say) x_1 odd, and $s \geqq \alpha + 1$, we can replace x_1 by $x_1 + 2^{s-\alpha}h$ and x_2 by $x_2 + 2^{s-\alpha}k$, and obtain $f \equiv n \pmod{2^{s+1}}$ if $2^sQ + 2^s(hx_2 + kx_1) \equiv 0 \pmod{2^{s+1}}$, where Q denotes an integer; i. e. with arbitrary h and unique k mod 2; thus again there are 2^{m-1} times as many solutions of $f \equiv n \pmod{2^{s+1}}$ as of $f \equiv n \pmod{2^s}$, with *any* x_i odd; and $\chi(2) \geqq 2^{-(m-1)(\alpha+1)}$.

This completes the proof of Theorem 4, hence of Theorem 1.

5. The criteria for *p*-adic universality and representation of zero. For any prime p we can number the variables so that (13) holds, with $m = 4$. Detailed proofs of the following three lemmas will be found in a forthcoming book by the author. Since their verification is not difficult we leave it to the reader, noting merely that the easiest proof of Lemma 2 amounts to a direct application of the conditions of Hasse [12] for p-adic representation of zero in rationals x_i.

LEMMA 1. *If $p > 2$, f is p-adically universal if and only if*

$$\alpha_1 = \alpha_2 = 0; \text{ and either } \alpha_3 = 0 \text{ or } (-a_1a_2 \mid p) = 1 \text{ or } \alpha_3 = \alpha_4 = 1. \tag{14}$$

If $p = 2$, the necessary and sufficient condition is:

$$\alpha_1 = 0,\ \alpha_2 = 0 \text{ or } 1,\ \alpha_3 - \alpha_2 = 0 \text{ or } 1; \tag{15}$$

$$\text{if } \alpha_2 = 0, \text{ then either } \alpha_4 - \alpha_3 \leqq 2 \text{ or } c^*_2 = 1; \tag{16}$$

$$\text{if } \alpha_2 = 1, \text{ then either } \alpha_4 - \alpha_3 \leqq 1 \text{ or } c^*_2 = 1. \tag{16}$$

Here c^*_2 denotes the unit

$$c^*_2 = (2 \mid a_1a_2)^{a_3}(2 \mid a_1a_3)^{a_2}(-1)^{\frac{1}{2}(a_1a_2+1)\cdot\frac{1}{2}(a_1a_3+1)}.$$

We may observe that c^*_2 is the invariant Ψ of Smith [13], or the invariant c_2 of Hasse, for the ternary form $g = a_1x_1^2 + 2^{\alpha_2}a_2x_2^2 + 2^{\alpha_3}a_3x_3^2$. It has the property that if $c^*_2 = 1$, then g represents zero 2-adically, and f does likewise.

LEMMA 2. *If $p > 2$ and (14) holds, then f fails to represent zero p-adically, if and only if*

(17) $\quad \alpha_1 = \alpha_2 = 0,\ \alpha_3 = \alpha_4 = 1,\ (-a_1a_2|p) = -1 = (-a_3a_4|p).$

If $p = 2$ and (15) holds, then f fails to represent zero p-adically if and only if $a_1a_2a_3a_4 \equiv 1 \pmod 8$ and any of the following cases (18)-(23) holds:

(18) $\quad \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0,\quad a_1 \equiv a_2 \equiv a_3 \equiv a_4 \pmod 4;$

(19) $\quad \alpha_1 = \alpha_2 = 0,\ \alpha_3 = \alpha_4 = 1,$ *and*
either (i) $a_1 \equiv a_2,\ a_3 \equiv a_4 \pmod 8,\ a_1 \equiv a_3 \pmod 4,$
or (ii) $a_1 \equiv 5a_2,\ a_3 \equiv 5a_4 \pmod 8,\ a_1 \equiv -a_3 \pmod 4,$
or (iii) $a_1 \equiv 3a_2,\ a_3 \equiv 3a_4 \pmod 8;$

(20) $\quad \alpha_1 = \alpha_2 = \alpha_3 = 0,\ \alpha_4 = 2,\ a_1 \equiv a_2 \equiv a_3 \equiv a_4 \pmod 4;$

(21) $\quad \alpha_1 = \alpha_2 = 0,\ \alpha_3 = 1,\ \alpha_4 = 3,$ *and three cases as in* (19);

(22) $\quad \alpha_1 = 0,\ \alpha_2 = \alpha_3 = 1,\ \alpha_4 = 2,$ *and the three cases of* (19) *with subscripts in the order* 1423;

(23) $\quad \alpha_1 = 0,\ \alpha_2 = 1,\ \alpha_3 = 2,\ \alpha_4 = 3,$ *and the three cases of* (19) *with subscripts in the order* 1324.

The least index s for which, in the cases of Lemma 2, $f \equiv 0 \pmod{p^s}$ implies that all x_i are divisible by p, is given as follows.

LEMMA 3. *When (17) holds, $f \equiv 0 \pmod{p^2}$ implies that p divides every x_i. If $p = 2$ and $a_1a_2a_3a_4 \equiv 1 \pmod 8$, then $f \equiv 0 \pmod{2^s}$ implies every x_i even, as follows: when $s = 3$ in (18), when $s = 4$ in (19), when $s = 5$ in (20), when $s = 6$ in (21)-(23).*

6. The forms which fail to represent zero p-adically for an odd p, and yet represent all large integers. By Lemma 3, if f does not represent n it does not represent $p^{2k}n$. Hence such forms f must represent all integers. By (17), f has the form

$$a_1x_1^2 + a_2x_2^2 + p(a_3x_3^2 + a_4x_4^2), \qquad (-a_1a_2|p) = -1 = (-a_3a_4|p),$$

and we can suppose $a_1 \leqq a_2$, $a_3 \leqq a_4$, $p \geqq 3$. In order that f shall represent 1 and 2, $a_1 = 1$ and $a_2 = 1$ or 2. If $p \geqq 7$, f cannot represent 6. If $p = 3$, $(-a_1a_2|p) = -1$ implies that $a_2 = 1$; then if f represents 3,

$a_3 = 1$ and $a_4 \equiv 1 \pmod 3$; if f represents 6, $a_4 = 1$. There remains form (A) of Theorem 3, which is known to represent all positive integers. If $p = 5$ then $a_2 = 2$ in order that $(-a_1a_2|p) = -1$; $a_3 = 1$ if f represents 5, and $a_4 \equiv 2$ or $3 \pmod 5$; $a_4 = 2$ if f represents 10. We thus obtain the form (B), which was overlooked by Kloosterman, and represents all positive integers.

7. The forms satisfying (18) which represent all large integers. By Lemma 3 such forms must represent all even positive integers. Let

$$f = (a_1, a_2, a_3, a_4), \quad a_1 \equiv a_2 \equiv a_3 \equiv a_4 \pmod 4), \quad a_1 \cdots a_4 \equiv 1 \pmod 8),$$
$$a_1 \leqq a_2 \leqq a_3 \leqq a_4.$$

If $a_1 > 1$, or if $a_1 = 1$ and $a_2 \geqq 5$, then $f \neq 2$. If $a_1 = a_2 = 1$ and $a_3 \geqq 9$, $f \neq 6$. If $a_1 = a_2 = 1$, $a_3 = 5$, $a_4 \geqq 9$, then $f \neq 12$. If $a_1 = a_2 = a_3 = 1$ and $a_4 \geqq 33$, then $f \neq 28$. There remain forms (C), treated in **10.**

8. The forms of type (20). We need only consider the forms (C) with a coefficient multiplied by 4, thus getting the nine distinct forms (D). By Lemma 3 it suffices to prove that these represent all multiples of 8. Since the forms (C) represent all large integers, their products by 4 represent all large multiples of 4. Hence the forms (D) represent all large multiples of 4; but none of these forms can fail to represent a number $8n$, since it would then not represent the large numbers $4^k \cdot 8n$. Hence the forms (D) represent all large integers.

9. Further cases in Lemma 2. Next, consider case (19), (i). We use ? to indicate that no multiple of 4 less than 200 is not represented:

$(1, 1, 2, 2t)$? ; $(1, 1, 2, \geqq 66) \neq 56$; $(1, 1, 10, 10)$? ;
$(1, 1, 10, \geqq 26) \neq 24$; $(1, 1, \geqq 18, \geqq 18) \neq 12$; $(1, 9, 2, 2t)$? ;
$(1, 9, 2, \geqq 66) \neq 56$; $(1, \geqq 9, \geqq 10, \geqq 10) \neq 8$;
$(1, 17, 2, 2s)$? if $s = 1, 9, 17$; $(1, 17, 2, \geqq 50) \neq 40$;
$(1, 25, 2, 2r)$? if $r = 1, 9$; $(1, 25, 2, \geqq 34) \neq 20$;
$(1, \geqq 33, 2, \geqq 2) \neq 28$; $(\geqq 3, \geqq 3, \geqq 2, \geqq 6) \neq 4$;
$(5, 5, 2, 2)$? ; $(\geqq 5, > 5, 2, 2) \neq 12$.

There remain the 15 forms (E), treated in **10-12.**

Case (21) (i) is got by multiplying one of the even coefficients in the preceding case by 4, yielding the 24 distinct forms (F). To represent all large numbers they need only represent all multiples of 16, and the fact that they do so follows, much as in **8**, from their connection with (E).

Cases (i) of (22) and (23) lead similarly to the $24+39$ forms (G) and (H), which represent all large numbers.

Next, take the case of (19) (ii):

$$(1, \geqq 5, \geqq 6, \geqq 14) \neq 8; \qquad (\geqq 3, \geqq 7, \geqq 2, \geqq 10) \neq 4.$$

Hence no such form represents all large integers. The same conclusion follows for cases (ii) of (21)-(23).

Lastly, consider cases (iii). By (19) we have: the nine forms (I) to be treated in **10-13**;

$$(1, u, 2, \geqq 54) \neq 40; \quad (1, \geqq 3, \geqq 6, \geqq 10) \neq 8;$$
$$(1, \geqq 27, 2, \geqq 6) \neq 20; \quad (\geqq 3, \geqq 3, \geqq 2, \geqq 6) \neq 4.$$

Extending these to (21)-(23) we get the 72 forms (J), (K), and (L).

10. The method of ternary sections. To complete the proof of Theorem 3 we need only prove that the forms in (A), (B), (C), (E), (I) represent all large integers. The investigation of such problems is usually made to depend on a knowledge of the numbers represented by a ternary section, obtained by putting one of the variables equal to zero. We know the numbers represented by a genus of ternary quadratic forms, and (with a few exceptions) it is only when a ternary form belongs to a genus of one class that we can tell precisely what numbers it represents. For example, the forms $(1,1,3)$, $(1,2,5)$, $(1,1,5)$, $(1,1,1)$, $(1,1,2)$, $(1,2,2)$, $(1,2,3)$, $(1,2,6)$, (needed in (A), (B), (C), (E), and (I)), are in genera of one class, and so are known to represent all positive integers except those of the respective forms $3^{2k}(9q+6)$, $5^{2k}(25q+10 \text{ or } 15)$, $4^k(8q+3)$, $4^k(8q+7)$, $4^k(16q+14)$, $4^k(8q+7)$, $4^k(16q+10)$, and $4^k(8q+5)$. The forms $(1,1,10)$, $(2,2,5)$, $(1,2,9)$, and $(1,2,18)$ (needed in (E)) are not in genera of one class. However, $(1,1,10)$ represents $2n$ if $(1,1,5)$ represents n; $(1,2,9)$ represents $2n$ if $g_0 = y^2 + 2x^2 - 2xz + 5x^2$ (in a genus of one class, representing all $\neq 4^k(8q+7)$) represents n; $(2,2,5)$ represents $4n$ if $(1,1,5)$ represents n; $(1,2,18)$ represents $4n$ if g_0 represents n. Only the forms in (2) and $(1,2,11,22)$ fail to succumb, by means of these facts, to the following treatment which we illustrate by the form $(1,25,2,18)$.

The form $(1,25,2,18)$ must be shown to represent all multiples of 4. Now $(1,2,18)$ represents all multiples of 4 except $4^{k+1}(8q+7)$. Also, $4^{k+1}(8q+7) - 4^{k+1}\cdot 25 = 4^{k+1}(8q-18)$, and this is represented by $(1,2,18)$ unless $q=0$, 1, or 2. Finally, $(1,25,2,18)$ is verified as representing 28, 60, and 92.

11. The forms (1, 2, 11, 22), (1, 2, 11, 38), and (1, 2, 19, 22). The form $(1, 2, 11)$ represents $2n$ if $g_1 = y^2 + 2x^2 - 2xz + 6z^2$ represents n. Now g_1 is in a genus of one class and represents all positive integers not of the form $4^k(8q + 5)$. Hence $(1, 2, 11)$ represents all evens $\neq 4^k(16q + 10)$. Also, $4^{k+1}(16q + 10) - 22 \cdot 4^k = 4^k(64q + 18)$, which is represented by $(1, 2, 11)$. Again, $4^{k+1}(16q + 10) - 38 \cdot 4^k = 4^k(64q + 2)$, also represented by $(1, 2, 11)$. Hence the forms $(1, 2, 11, 22)$ and $(1, 2, 11, 38)$ represent all large multiples of 4; that they represent all large numbers not divisible by 4 follows from Theorem 4. Similarly, $(1, 2, 22)$ represents every $4n \neq 4^{k+1}(8q + 5)$; hence $(1, 2, 22, 19)$ represents every $4n$.

12. The form (1, 2, 17, 34). The form $x^2 + 2y^2 + 17z^2$ represents $2n$ if $g_2 = y^2 + 2x^2 - 2xz + 9z^2$ represents n. Now g_2 is not in a genus of one class, but is in a genus of two classes, the other class containing the form $h_2 = x^2 + y^2 + 17z^2$. Together, g_2 and h_2 represent all positive integers $\neq 4^k(8q + 7)$. However, the identity

$$(24) \quad x^2 + y^2 + 17z^2 = x^2 + 2(3z)^2 - 2(3z)(y + z)/3 + 9[(y + z)/3]^2$$

shows that if n is represented in h_2 with either $y + z$ or $x + z$ divisible by 3, then n is represented also in g_2. This can be arranged unless $x^2 \equiv y^2 \not\equiv z^2 \pmod 3$, whence $n \equiv 2 \pmod 3$. Thus:

g_2 represents every $3s$ and $3s + 1$ not of the form $4^k(8q + 7)$.

Hence $(1, 2, 17)$ represents every $6s$ and $6s + 2$ not of the form $4^k(16q + 14)$. In view of Theorem 4 we need consider only multiples of 4. If $6s$ or $6s + 2 = 4^{k+1}(16q + 14)$, then $6s - 34 \cdot 4^k = 4^k(64q + 22) \equiv 2 \pmod 6$, or $6s + 2 - 34 \cdot 9 \cdot 4^k = 4^k(64q - 250) \equiv 2 \pmod 6$, and thus is represented in $(1, 2, 17)$; except, in the last case when $q = 0$ or 3. But $(1, 2, 17, 34)$ represents 56 and 248.

Finally, there remains $6s + 4 = 4(3s_1 + 1)$, say. Subtracting 34 we get $12s_1 - 30$, which has the form $4^k(16q + 14)$ only if $s_1 = 4s_2 + 1$, whence $4(3s_1 + 1) = 4^2(3s_2 + 1)$. We proceed by induction. If $4^h(3s_h + 1) - 34 \cdot 2^{2h-2}$ or $4^{h-1}(12s_h - 30)$ is of the excluded form, then $s_h = 4s_{h+1} + 1$; and so on. If $4^{h-1}(12s_h - 30)$ is negative, s_h is 0, 1, or 2, and $4^h(3s_h + 1)$ is 4^h, 4^{h+1}, or $4^h \cdot 7$, all of which are evidently represented by $(1, 2, 17)$ if $h \geqq 1$.

13. The form (1, 2, 19, 38). The form $(1, 2, 19)$ represents $2n$ if $g_3 = x^2 + 2y^2 + 2yz + 10z^2$ represents n. Also, g_3 and $h_2 = 2x^2 + 2y^2 + 7z^2 + 2yz + 2zx + 2xy$ constitute a genus, representing all positive integers

$\neq 4^k(8q+5)$. This time, transformations of denominator 3, as in (24), do not suffice, and we use denominator 5. The transformations with the matrices $T/5$, where

$$T = \begin{pmatrix} 6 & 4 & -3 \\ 1 & 4 & 7 \\ 1 & -1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 6 & 2 & 9 \\ 1 & -3 & -6 \\ 1 & 2 & -1 \end{pmatrix}, \quad \begin{pmatrix} 4 & -2 & 7 \\ -4 & -3 & 3 \\ 1 & 2 & 3 \end{pmatrix},$$

replace g_3 by h_3. Hence if $n = 2x^2 + 2y^2 + 7z^2 + 2yz + 2zx + 2xy$ in integers x, y, z, then g_3 also represents n if any of the following congruences holds:

$$x - y + 2z \equiv 0, \quad x + 2y - z \equiv 0, \quad x + 2y - 2z \equiv 0,$$
$$x - y - 2z \equiv 0, \quad 2x + y - z \equiv 0, \quad 2x + y - 2z \equiv 0, \pmod 5,$$

the last three being got by interchanging x and y. Introducing $X = y + z$, $Y = z + x$, $Z = x + y$, we see that these conditions reduce to

$$(25) \qquad 2X \equiv Y \text{ or } Z, \text{ or } 2Y \equiv Z \text{ or } X, \text{ or } 2Z \equiv X \text{ or } Y \pmod 5.$$

If $n \equiv 0$, 1, or 4 (mod 5), then $h_3 = n$ implies $X^2 + Y^2 + Z^2 \equiv 0$, 1, or 4, whence (X, Y, Z) is a permutation of the following residues mod 5: $(0, 0, 0)$, $(0, 0, \pm 1)$, $(0, 0, \pm 2)$, $(0, \pm 1, \pm 2)$, $(\pm 1, \pm 1, \pm 2)$, $(\pm 1, \pm 2, \pm 2)$. In all cases, (25) is seen to hold (e. g. $2 \cdot 2 \equiv -1$, $2 \cdot 1 \equiv 2$). Hence:

g_3 represents every $5s + 0, 1, 4$ not of the form $4^k(8q+5)$;

and $(1, 2, 19)$ represents every $10s + 0, 2, 8$ not of the form $4^k(16q + 10)$.

If $10s = 4^{k+1}(16q + 10)$, then $10s - 38 \cdot 4^k = 4^k(64q + 2)$ and is represented by $(1, 2, 19)$. If $10s + 2 = 4^{k+1}(16q + 10)$, then according as k is odd or even, $10s + 2 - 38 \cdot 4^k = 4^k(64q + 2) \equiv 0 \pmod{10}$, or $10s + 2 - 38 \cdot 4^{k+1} = 4^{k+2}(4q - 7) \equiv 0 \pmod{10}$; in the last case q cannot be 0 or 1. Similar results hold if $10s + 8 = 4^{k+1}(16q + 10)$ with k even or odd.

Finally, we have $4(5s_1 \pm 1)$ to consider. Now $4(5s_1 \pm 1) - 38 = 4^k(16q + 10)$ only if $s_1 \equiv \mp 1 \pmod 4$ respectively. Then $4(5s_1 \pm 1) = 4^2(5s_2 \mp 1)$. We proceed by induction and have $4^h(5s_h \pm 1) - 38 \cdot 4^{h-1} = 4^{h-1}(20s_h \pm 4 - 38)$ of the form $4^k(16q + 10)$ only if $s_h = 4s_{h+1} \mp 1$; and so on. If $20s_h \pm 4 - 38$ is negative, then s_h is 0, 1, or 2; and we see that $(1, 2, 19, 38)$ represents 4, 16, 24, 36, and 44.

McGill University.

REFERENCES.

1. J. Liouville, *Journal de Mathématiques*, Ser. 2, vol. 1 (1856), p. 230.

2. S. Ramanujan, *Proceedings of the Cambridge Philosophical Society*, vol. 19 (1917), pp. 11-21; *Collected Papers*, 169-178.

3. L. E. Dickson, *Bulletin of the American Mathematical Society*, vol. 33 (1927), pp. 63-70; Dickson gives proofs of Ramanujan's results on ternary forms.

4. P. R. Halmos, *Bulletin of the American Mathematical Society*, vol. 44 (1938), pp. 141-144. See also A. E. Ross, *Bulletin of the American Mathematical Society*, vol. 49 (1943), p. 362.

5. H. D. Kloosterman, *Acta Mathematica*, vol. 49 (1926), pp. 407-464.

6. Compare G. Pall, *Bulletin of the American Mathematical Society*, vol. 46 (1940), p. 291.

7. Kloosterman, *loc. cit.*, p. 409.

8. C. L. Siegel, *Annals of Mathematics*, vol. 36 (1935), pp. 527-606.

9. W. Tartakowsky, *Bulletin of the Russian Academy of Sciences*, (7), 2 (1929), pp. 111-122.

10. Kloosterman, *loc. cit.*, p. 449.

11. This is of course merely an adaptation of the well-known stepping-up process. Compare Dickson's *Introduction to the Theory of Numbers*, p. 13.

12. H. Hasse, *Journal für Mathematik*, vol. 152 (1923), pp. 129-148 (Theorems 10, 11).

13. H. J. S. Smith, *Collected Mathematical Papers*, I, p. 464.

AN EXTENSION OF A PROBLEM OF KLOOSTERMAN.*

By Arnold E. Ross and Gordon Pall.

1. Introduction. In the previous article [1],[1] Kloosterman's problem of determining all the positive quaternary forms $ax^2 + by^2 + cz^2 + dt^2$ which represent all large integers was completely solved. Ross [2] has proved the following lemma, which makes it possible to extend the solution to cover the general positive integral quaternary form $\Sigma a_{ij}x_ix_j$ in which the a_{ii} and $2a_{ij}$ $(i, j = 1, \cdots, 4)$ are integers.

Lemma 1. *The determinant of any positive integral quaternary quadratic form which represents all positive integers cannot exceed a certain constant R_1.*

We make use also of an extension to a general positive m-ary form $(m \geqq 4)$ of Kloosterman's asymptotic formula [3] for the number of representations of n. Such a formula has been given by W. Tartakowsky [4], who used the Hardy-Littlewood method. It is interesting to see (2) how easily this extension can be made directly from Kloosterman's special case, by induction, making use of Ross's [5] technique of employing primes represented by a primitive form, and the fact that every primitive form in two or more variables represents infinitely many primes.

References to the preceding article will be prefixed by P. The definitions of the terms "f is universal for p," and "f represents zero p-adically," given near $P(3)$ and $P(4)$ evidently extend to any integral forms f. Once the asymptotic formula (1) is established, Lemma 2 and the proof of Theorem $P1$ yield

Theorem 1. *Let f be any positive integral m-ary quadratic form, $m \geqq 4$. Let n be such that $f(x_1, \cdots, x_m) \equiv n \bmod p^r$ is solvable for every p and r. For each prime p (if any) such that f fails to represent zero p-adically impose an upper bound to the power of p which may divide n. Then f represents every such n sufficiently large.*

It should be observed that if $m \geqq 5$, every f represents zero p-adically for every $p \mid 6$. Hence we can conclude (as Tartakowsky did) that all forms in a

* Received July 17, 1945.

[1] Numbers in square brackets refer to the references at the end of the paper.

genus of positive quadratic forms in five or more variables represent the same sufficiently large numbers. If $m = 4$, f can fail to represent zero p-adically only for a few primes appearing in its determinant (the conditions [6] are easily applied and are given partly in our Lemma 4).

It follows that if f is universal for every p, and represents zero p-adically for every p, then if $m \geqq 4$, f represents every sufficiently large positive integer n. In **4** we shall, using Lemma 1, obtain

THEOREM 2. *The number of integral positive quaternary quadratic forms which represent all large numbers, and yet fail to represent zero p-adically for some prime p_1, is finite. In fact the determinant of any such form cannot exceed R_1 if $p_1 > 2$, $256R_1$ if $p_1 = 2$.*

Theorem 1 does not extend to $m = 2$, since binary classes in the same genus, which are neither properly nor improperly equivalent, do not represent the same primes 7. The theorem does not extend without modification to $m = 3$, as was shown by an example in 1939 [8]. The example consists of the forms

$$f = x^2 + y^2 + 16z^2, \qquad g = 2x^2 + 2y^2 + 5z^2 - 2xz - 2yz,$$

which are easily seen to represent equally often every positive integer not an odd square. If s is odd, a result of Jacobi's [9] shows that

$$f(s^2) - g(s^2) = (-1)^{\frac{1}{2}(s-1)} 4s,$$

where $f(n)$ denotes the number of representations of n by f. Also,

$$f(s^2) + g(s^2) = 4\Pi\psi(p, a),$$

where

$$\psi(p, a) = (p^{a+1} - 1)/(p - 1) - (-1|p)(p^a - 1)/(p - 1),$$

and the product ranges over the prime-power decomposition $s = \Pi p^a$. From this it is easily seen that $g(s^2) = 0$ if every prime p in s satisfies $p \equiv 1 \bmod 4$. This example is especially interesting in that f and g are two forms in the same genus, and we are able to give exact simple formulas for the number of representations of any number in either form.

However, the theorem is true when $m = 3$ for a large variety of forms [10]; and, if we may make a conjecture, it is probably true for squarefree numbers n and every ternary f.

2. The asymptotic formula for the number of representations of *n* in *f*.

THEOREM 3. *Let $m \geqq 4$. For any positive integral m-ary form $f = \Sigma a_{ij} x_i x_j$ where a_{ii} and $2a_{ij}$ $(i, j = 1, \cdots, m)$ are integers, the number of representations $f(n)$ of n by f is given by the formula*

$$\frac{\pi^{\frac{1}{2}m} n^{\frac{1}{2}m-1} S(n)}{\Gamma(\frac{1}{2}m)\Delta^{\frac{1}{2}}} + O(n^{\frac{1}{2}m-1-1/18+\epsilon}), \tag{1}$$

where Δ denotes the determinant $|a_{ij}|$ of f,

$$S(n) = \prod_{p=2,3,5,7,\ldots} \chi(p), \qquad \chi(p) = \chi(p, f, n) = \lim_{r\to\infty} p^{-(m-1)r} f(n, p^r), \tag{2}$$

and $f(n, p^r)$ denotes the number of solutions $x_1, \cdots, x_m$ mod p^r of $f \equiv n \bmod p^r$.

Proof. We can write $f = b_1 x_1^2 + \cdots + b_{s-1} x_{s-1}^2 + d_1 \phi(x_s, \cdots, x_m)$, where ϕ is primitive. Let Δ_1 denote the determinant of ϕ; then $\Delta = b_1 \cdots b_{s-1} d_1^{m-s+1} \Delta_1$. Suppose, then, that (1) is known to be true if $s_1 \leqq s \leqq m$. Kloosterman's case is of course $s_1 = m$. Now let $s = s_1 - 1$. Let q denote an odd prime represented by ϕ and not dividing Δ. We can write $\phi = \sum_{i,j=s}^{m} a_{ij} x_i x_j$, $a_{ss} = q$.

Case I: the a_{sj} all integers. By completing squares we have the identity

$$qf = qb_1 x_1^2 + \cdots + qb_{s-1} x_{s-1}^2 + d_1 (q x_s + \sum_{j=s+1}^{m} a_{sj} x_j)^2 + d_1 \psi, \tag{3}$$

where $\psi(x_{s+1}, \cdots, x_m) = \sum_{i,j=s+1}^{m} (q a_{ij} - a_{is} a_{sj}) x_i x_j$. We introduce two forms:

$$f_1 = qb_1 y_1^2 + \cdots + qb_{s-1} y_{s-1}^2 + d_1 y_s^2 + d_1 \psi(y_{s+1}, \cdots, y_m), \tag{4}$$

and the form f_2 got from f_1 by changing the coefficient of y_s^2 to $q^2 d_1$. Now for every representation of qn in f_1 we have

$$y_s^2 \equiv -\psi(y_{s+1}, \cdots, y_m) \equiv \sum_{i,j=s+1}^{m} a_{is} y_i \cdot a_{sj} y_j \equiv (\sum_{j=s+1}^{m} a_{sj} y_j)^2 \pmod q.$$

Hence: $y_s \equiv \Sigma a_{sj} y_j \bmod q$ holds only by choice of sign of y_s if $q \nmid y_s$, and always if $q \mid y_s$. Accordingly,

$$\begin{aligned} f(n) &= N(n = f) = N(qn = qf) = N(qn = f_1;\ y_s \equiv \Sigma a_{sj} y_j \bmod q) \\ &= \tfrac{1}{2}\{N(qn = f_1) - N(qn = f_2)\} + N(qn = f_2), \text{ that is,} \end{aligned}$$

$$f(n) = \tfrac{1}{2} f_1(qn) + \tfrac{1}{2} f_2(qn). \tag{5}$$

If $p \neq q$, f_1 and f_2 are derived from qf by transformations of determinants prime to p, whence $f(n, p^r) = f_1(qn, p^r) = f_2(qn, p^r)$. Hence

$$(6) \qquad \chi(p, f, n) = \chi(p, f_1, qn) = \chi(p, f_2, qn), \text{ if } p \neq q.$$

If $p = q$, we have (counting solutions of the congruences with care):

$$\begin{aligned} f(n, q^r) = q^{-m} f(qn, q^{r+1}) &= q^{-m+1} N(qn \equiv f_1 \bmod q^{r+1};\ y_s \equiv \Sigma a_{sj} y_j \bmod q) \\ &= q^{-m+1} \tfrac{1}{2} \{ N(qn \equiv f_1 \bmod q^{r+1}) - q^{-1} N(qn \equiv f_2 \bmod q^{r+1}) \} \\ &\quad + q^{-m+1} q^{-1} N(qn \equiv f_2 \bmod q^{r+1}), \end{aligned}$$

whence

$$(7) \qquad f(n, q^r) = q^{-m+1} \{ \tfrac{1}{2} f_1(qn, q^{r+1}) + \tfrac{1}{2} q^{-1} f_2(qn, q^{r+1}) \},$$

$$(8) \qquad \chi(q, f, n) = \tfrac{1}{2} \chi(q, f_1, qn) + \tfrac{1}{2} q^{-1} \chi(q, f_2, qn).$$

Hence as (1) holds for f_1 and f_2, we have by (5),

$$f(n) = \{ \Gamma(\tfrac{1}{2} m) \}^{-1} \pi^{\frac{1}{2} m} \Delta^{-\frac{1}{2}} n^{\frac{1}{2} m - 1} \prod_{p \neq q} \chi(p) \cdot \lambda + O((qn)^{\frac{1}{2} m - 1 - 1/18 + \epsilon}),$$

where

$$\lambda = \tfrac{1}{2} (q^{m-2})^{-\frac{1}{2}} q^{\frac{1}{2} m - 1} \chi(q, f_1, qn) + \tfrac{1}{2} (q^m)^{-\frac{1}{2}} q^{\frac{1}{2} m - 1} \chi(q, f_2, qn) = \chi(q, f, n),$$

by (8). The induction is thus complete for Case I.

Case II: some coefficient $2a_{sj}$ is odd. We now use the identity

$$4qf = 4qb_1 x_1^2 + \cdots + 4qb_{s-1} x_{s-1}^2 + d_1 (2qx_s + \Sigma 2a_{sj} x_j)^2 + d_1 \psi,$$

where $\psi = \Sigma (4qa_{ij} - 4a_{is} a_{sj}) x_i x_j$. Besides the two forms

$$(9) \quad f_k = 4qb_1 y_1^2 + \cdots + 4qb_{s-1} y_{s-1}^2 + q^{2k-2} d_1 y_s^2 + d_1 \psi(y_{s+1}, \cdots, y_m), \qquad k = 1, 2,$$

we shall require the form f'_k obtained from f_k by changing y_s to $2y_s$, the form f_k'' obtained from f_k by expressing the condition that $\Sigma 2a_{sj} y_j$ is even (e. g. if $2a_{sm}$ is odd by replacing y_m in ψ by $2a_{s,s+1} y_{s+1} + \cdots + 2a_{s,m-1} y_{m-1} + 2y_m$), and the form f_k''' obtained by both the preceding operations.

If $4qn$ is represented in f_1, we now have only $y_s^2 \equiv (\Sigma 2a_{sj} y_j)^2 \bmod q$. Hence for any integer n,

$$\begin{aligned} f(n) = N(4qn = 4qf) &= N(4qn = f_1;\ y_s \equiv \Sigma 2a_{sj} y_j \bmod 2q) \\ &= \tfrac{1}{2} N(4qn = f_1, y_s \equiv \Sigma 2a_{sj} y_j \bmod 2) + \tfrac{1}{2} N(4qn = f_2, y_s \equiv \Sigma 2a_{sj} y_j \bmod 2). \end{aligned}$$

Hence it is easily seen that

(10) $$f(n) = \sum_{k=1}^{2} \{\tfrac{1}{2}f_k(4qn) - \tfrac{1}{2}f_k'(4qn) - \tfrac{1}{2}f_k''(4qn) + f_k(4qn)\}.$$

If $p \neq q$ and $p \neq 2$, we obviously have

$$f(n, p^r) = f_1(4qn, p^r) = f_1'(4qn, p^r) = \cdots = f_2'''(4qn, p^r),$$

whence $\chi(p)$ is the same in all cases. Evidently also $\chi(2)$ is the same for f_1 as for f_2, for f_1' as for f_2', for f_1'' as for f_2'', for f_1''' as for f_2'''; and $\chi(q)$ is the same for all four forms f_1, f_1', f_1'', f_1''', and again for the forms f_2, f_2', f''_2, f_2'''. Next,

$$f(n, q^r) = q^{-m}f(4qn, q^{r+1}) = q^{1-m}N(f_1 \equiv 4qn \bmod q^{r+1}; y_s \equiv \Sigma 2a_{sj}y_j \bmod q)$$
$$= q^{1-m}\{\tfrac{1}{2}f_1(4qn, q^{r+1}) + \tfrac{1}{2}q^{-1}f_2(4qn, q^{r+1})\};$$

(11) $$\chi(q, f, n) = \tfrac{1}{2}\chi(q, f_1, 4qn) + \tfrac{1}{2}q^{-1}\chi(q, f_2, 4qn).$$

Also, $f(n, 2^r) = 4^{-m}N(4qf \equiv 4qn \bmod 2^{r+2}) = 2^{1-2m}N(f_1 \equiv 4qn \bmod 2^{r+2}$; $y_s \equiv \Sigma 2a_{sj}y_j \bmod 2) = 2^{1-2m}f_1(4qn, 2^{r+2}) - N(f_1 \equiv 4qn$, y_s even$)$ $- N(f_1 \equiv 4qn$, $\Sigma 2a_{sj}y_j$ even$) + 2N(f_1 \equiv 4qn$, y_s and $\Sigma 2a_{sj}y_j$ even$)$ whence

(12) $$\chi(2, f, n) = \tfrac{1}{2}\chi(2, f_1, 4qn) - \tfrac{1}{4}\chi(2, f_1', 4qn) - \tfrac{1}{4}\chi(2, f_1'', 4qn)$$
$$+ \tfrac{1}{4}\chi(2, f_1''', 4qn).$$

Substituting in (10) we obtain the required formula for $f(n)$ in this case also.

3. Proof of Theorem 1. This proceeds exactly as in P **3**. We have only to convert f into a convenient form-residue mod p^r: cf. Lemma 2 [11]. Corresponding to $P(12)$, with Δ in place of $abcd$, we can use known formulas [12] for $f(n, p^r)$ in the cases where $p \nmid 2\Delta$ to prove that $\chi(p) \geqq 1 - p^{-\frac{1}{2}m}$ if $p \nmid n$ and m is even, $\chi(p) \geqq (1 - p^{-\frac{1}{2}m})(1 - p^{-1})$ if $p \mid n$ and m is even, and $\chi(p) \geqq 1 - p^{1-m}$ if m is odd. Hence the product of the $\chi(p)$ for all odd primes not dividing Δ is easily seen to exceed $K/\log\log n$. The additional discussion for cases (14)-(17) will be found in P, **3**.

Lemma 2. *Every integral quaternary form f is equivalent* mod p^r, *by a transformation of determinant prime to p, to a form of the type*

(13) $$p^{\alpha_1}a_1x_1^2 + \cdots + p^{\alpha_4}a_4x_4^2, \quad \bmod p^r,$$

$$0 \leqq \alpha_1 \leqq \alpha_2 \leqq \alpha_3 \leqq \alpha_4, \quad a_1a_2a_3a_4 \text{ prime to } p,$$

if $p > 2$. If $p = 2$, f is equivalent mod 2^r *either to* (13), *or to a form with one of the following residues* mod 2^r:

$$(14) \qquad 2^{\alpha_1}a_1x_1^2 + 2^{\alpha_2}a_2x_2^2 + 2^{\alpha_3}(2jx_3^2 + 2x_3x_4 + 2jx_4^2),$$

$$(15) \qquad 2^{\alpha_1}a_1x_1^2 + 2^{\alpha_2}(2jx_2^2 + 2x_2x_3 + 2jx_3^2) + 2^{\alpha_4}a_4x_4^2,$$

$$(16) \qquad 2^{\alpha_1}(jx_1^2 + x_1x_2 + jx_2^2) + 2^{\alpha_3}a_3x_3^2 + 2^{\alpha_4}a_4x_4^2,$$

$$(17) \qquad 2^{\alpha_1}(jx_1^2 + x_1x_2 + jx_2^2) + 2^{\alpha_3}(kx_3^2 + x_3x_4 + kx_4^2).$$

Here all a_i are odd, $j, k = 0$ or 1. Also, in (14), $0 \leqq \alpha_1 \leqq \alpha_2 < \alpha_3$; *in* (15), $0 \leqq \alpha_1 < \alpha_2 < \alpha_4$; *in* (16), $0 \leqq \alpha_1 \leqq \alpha_3 \leqq \alpha_4$; *and in* (17), $0 \leqq \alpha_1 \leqq \alpha_3$.

4. Proof of Theorem 2. We first extend Lemmas $P1$ and $P2$ as follows:

LEMMA 3. *The conditions that f be universal for p are as stated in Lemma P1 if f has the residue* (13). *If $p = 2$ and f has a residue* (14) *or* (15), *f is not universal for p. If $p = 2$ and f has a residue* (16) *or* (17), *f is universal for p if and only if, respectively:*

$$(18) \qquad \alpha_1 = 0, \text{ and either } j = 0 \text{ or } \alpha_3 = 0 \text{ or } \alpha_3 = 1 \geqq \alpha_4 - 2;$$

$$(19) \qquad \alpha_1 = 0, \text{ and either } j = 0 \text{ or } \alpha_3 = 0 \text{ or } \alpha_3 = 1.$$

LEMMA 4. *If f is universal for p, and f has a residue* (13), *the condition that f fail to represent zero p-adically is given in Lemma P2. If* (16) *and* (18), *or* (17) *and* (19), *hold, then f fails to represent zero p-adically if and only if, respectively:*

$$(20) \qquad \alpha_1 = 0,\ j = 1,\ \alpha_3 = 1,\ \alpha_4 = 1 \text{ or } 3, \text{ and } a_3a_4 \equiv 3 \bmod 8;$$

$$(21) \qquad \alpha_1 = 0,\ j = 1,\ \alpha_3 = 1,\ k = 1.$$

Again, Lemma $P3$ extends to the cases coming under (13). Also,

LEMMA 5. *$2^s | f$ implies $2 |$ all x_i, if $s = 3 + \alpha_4$ in* (20), *if $s = 2$ in* (21).

If $p > 2$, or if $p = 2$ and the cases corresponding to (21) hold, then if f does not represent one number n, f will not represent $p^{2k}n$, by Lemmas 5 and $P3$. Hence if f represents all large integers, $\det f \leqq R_1$.

Let $p = 2$ and f be given by (13). (a) If $P(18)$ holds and f represents all large integers it must represent every $2n$, by Lemma $P3$. Now if $f = 2n$, Σx_i is even. Hence the transformation $x_1 = y_1$, $x_2 = y_2$, $x_3 = y_3$, $x_4 = y_1 + y_2 + y_3 + 2y_4$ replaces f by $2g$, where g is an integral form and g represents every n; by Lemma 1, $\det g \leqq R_1$, $\det f \leqq 4R_1$. (b) If $P(20)$ holds, f is

obtainable (by changing x_4 to $2x_4$) from a form under case (a). Hence $\det f \leqq 16R_1$. (c) If $P(19)$ holds, and f represents all large integers, f represents every $4n$. In subcases (i) and (ii), $f = 4n$ implies $x_1 = y_1 + y_2$, $x_2 = y_1 - y_2$, $x_3 = y_3$, $x_4 = y_1 + y_2 + y_3 + 2y_4$ where the y_i are integers; and f becomes $4g$ where g represents all n; hence $\det f \leqq 16R_1$. In subcase (iii), $x_1 = y_1 + y_2$, $x_2 = y_1 - y_2$, $x_3 = y_3 + y_4$, $x_4 = y_3 - y_4$ yields the same result. (d) As in case (b), forms coming under $P(21)$-(23) lead to $\det f \leqq 64R_1$, $64R_1$, $256R_1$ respectively.

Next let $p = 2$ and f correspond to (20) with $\alpha_4 = 1$. Then f must represent every $4n$. If $f = 4n$, $x_1 = 2y_1$, $x_2 = 2y_2$, $x_3 = y_3 + y_4$, $x_4 = y_3 - y_4$, and we get $\det f \leqq 4R_1$. Finally, if $\alpha_4 = 3$ instead, $\det f \leqq 16R_1$.

ST. LOUIS UNIVERSITY,
MCGILL UNIVERSITY.

REFERENCES.

1. G. Pall, *American Journal of Mathematics*, vol. 68 (1946), pp. 47-58.
2. A. E. Ross, *American Journal of Mathematics*, vol. 68 (1946), pp. 29-46.
3. H. D. Kloosterman, *Acta Mathematica*, vol. 49 (1926), pp. 407-464.
4. W. Tartakowsky, *Bull. Acad. Sci. Leningrad* (7) 2 (1929), pp. 111-122, 165-196.
5. A. E. Ross, *loc. cit.*
6. H. Hasse, *Journal für Mathematik*, vol. 152 (1923), pp. 129-148.
7. E. g., G. Pall, *Mathematische Zeitschrift*, vol. 36 (1933), pp. 321-343, (329).
8. B. W. Jones and G. Pall, *Acta Mathematica*, vol. 70 (1939), pp. 165-191 (181).
9. Cf. Dickson's *History of the Theory of Numbers*, vol. II, pp. 261-263.
10. U. V. Linnik, *Bull. Acad. Sci. URSS. Ser. Math.*, vol. 4 (1940), pp. 363-402; a correction promised in *C. R. Acad. Sci. URSS.*, sometime in 1943.
11. Cf. B. W. Jones, *Duke Mathematical Journal*, vol. 9 (1942), pp. 723-756 (726-727).
12. E. g., C. L. Siegel, *Annals of Mathematics*, vol. 36 (1935), pp. 527-606, (Lemma 16).

GENERAL THEORY OF SCALE CURVES.*

By EDWARD KASNER and JOHN DE CICCO.

1. Cartograms and scale curves. Let (x, y) denote general curvilinear coordinates of a point on a surface Σ and cartesian coordinates on a plane Π. This defines a point transformation T between the points of Σ and Π such that points correspond if they are represented by the same coordinates. We shall term any particular mapping of a surface Σ upon a plane Π a *cartogram*. Thus a cartogram depends not only on the surface Σ but also on the point transformation T. It is said to be conformal or non-conformal according as T is conformal or non-conformal.

The *scale function* $\sigma = ds/dS$ is the ratio of the differentials of arc lengths of the corresponding curves on Π and Σ respectively, under the transformation T. In general, σ depends not only on the point (x, y) but also on the slope $y' = dy/dx$; that is, $\sigma = \sigma(x, y, y')$. It is independent of the direction if, and only if, T is conformal. The scale σ is a mere constant only in the degenerate situation where Σ is developable and the mapping T is an unrolling of Σ upon Π followed by a similitude in Π.

A *scale curve* is the locus of a point on Σ or Π along which the scale σ does not vary. Thus the totality of scale curves for any given cartogram, is defined by $\sigma =$ const. For a non-conformal cartogram, there are ∞^2 scale curves. For a conformal cartogram, there are ∞^1 scale curves, except in the degenerate situation mentioned above where every curve is a scale curve.

Since we have already studied the simple-infinity of scale curves in conformal maps, we shall assume in the present paper that the mapping is not conformal. Thus the slope y' is explicitly present in the scale function σ. We shall study the geometry of systems of ∞^2 scale curves in the plane Π for non-conformal cartograms.

2. General summary. Upon solving the differential equation of the second order defining the ∞^2 scale curves for a given non-conformal cartogram for the second derivative $y'' = d^2y/dx^2$, it is found that y'' is expressed as a rational function of y' of the fifth degree with coefficients depending on six

* Received July 3, 1945. Presented to the American Mathematical Society, February, 1944.

functions of (x, y). These six functions are not arbitrary but satisfy a system of two partial differential equations of the second order. (Theorem 3).

In general, there are essentially ∞^2 surfaces which possess the same family of ∞^2 curves as scale curves.

At any point of the plane Π, *there are three inflectional directions and two cuspidal directions for the system of ∞^2 scale curves.* (Theorem 1). The cuspidal directions are always orthogonal and coincide with the characteristic directions.

Through a point of the plane Π there pass ∞^1 scale curves. Upon constructing the osculating circles to these scale curves at the point, *the centers of curvature describe a general cubic curve (the central locus) which has a node at the fixed point, the tangent directions of the node coinciding with the characteristic directions.* (Theorem 2). Of course, this property is not completely characteristic for systems of scale curves.

A system of ∞^2 scale curves is of the cubic (Lie-Liouville) type if, and only if, it is a velocity system. (Theorem 5). In that event, the cubical locus degenerates into a straight line (the other two straight lines becoming the tangent lines of the characteristic directions). *If the scale curves form a natural family, then the characteristic curves form an isothermal net.* (Theorems 7 and 8).

We find *a new class of surfaces Σ for which there exists a map of Σ upon* Π *such that the scale curves coincide with the totality of ∞^2 straight lines of* Π. Corresponding to a given scale σ, there are ∞^1 straight scales all tangent to a conic section. As σ varies, the resulting ∞^1 conics form a confocal family.

There are essentially ∞^6 cartograms with straight lines as scales. The only such surfaces of constant gaussian curvature are developable surfaces. *Thus there exists no non-conformal map T of a sphere (or pseudo-sphere) Σ upon a plane* Π *such that the scale curves are all straight lines.* (Theorem 9). As a consequence of this, it follows that the only map, conformal or not, of a sphere upon a plane such that the scale curves are pictured by straight lines, is the Mercator projection, in which case there is only a simple-infinitude of straight scales.

3. The differential equation of the ∞^2 scale curves of a non-conformal cartogram. Let (x, y) denote general curvilinear coordinates of a point on a surface Σ and cartesian coordinates on a plane Π. The square of the linear element dS of Σ is

$$(1) \qquad dS^2 = E(x, y)\,dx^2 + 2F(x, y)\,dx\,dy + G(x, y)\,dy^2,$$

where $H^2 = EG - F^2 > 0$, and the square of the linear element ds of Π is

$ds^2 = dx^2 + dy^2$. This defines a point transformation T between the points of Σ and Π such that points correspond if they are represented by the same coordinates. Since we are working with non-conformal cartograms, it follows that the conditions $F = 0$ and $E = G$, can never hold simultaneously.

The gaussian curvature G of the surface Σ is given by

$$
\begin{aligned}
(2) \qquad 4H^4 G = {} & E(G_x{}^2 - 2F_x G_y + E_y G_y) + G(E_y{}^2 - 2E_x F_y + E_x G_x) \\
& + F(E_x G_y - E_y G_x - 2F_x G_x + 4F_x F_y - 2E_y F_y) \\
& - 2H^2(E_{yy} - 2F_{xy} + G_{xx}).
\end{aligned}
$$

This will be found useful for our later work.

The scale function $\sigma = ds/dS$ is the ratio of the differentials of arc lengths of the corresponding curves on Π and Σ, respectively, under the transformation T. It is defined by the formula

$$
(3) \qquad \sigma^2 = \left(\frac{ds}{dS}\right)^2 = \frac{1 + y'^2}{E + 2Fy' + Gy'^2}.
$$

Thus σ is a function of the lineal-element (x, y, y'). For a non-conformal cartogram, the slope y' is explicitly present in σ since the conditions $F = 0$ and $E = G$ do not hold simultaneously.

The scale curves are defined by $\sigma = \text{const.}$ *There are always ∞^2 scale curves in a non-conformal cartogram.* Upon eliminating the constant σ from the equation (3) by differentiation, we discover that the differential equation of the ∞^2 scale curves is

$$
(4) \qquad y'' = \frac{(1 + y'^2)[E_x + y'(E_y + 2F_x) + y'^2(2F_y + G_x) + y'^3 G_y]}{2[-F + y'(E - G) + y'^2 F]}.
$$

Not every system of ∞^2 curves in the plane Π can represent the scales of a non-conformal cartogram, since (4) is of special algebraic form in the first derivative.

THEOREM 1. *At a point, there are three inflectional directions and two cuspidal directions for the scale curves. The cuspidal directions are always orthogonal and coincide with the characteristic directions.*

This may be deduced from the differential equation (4) of the ∞^2 scale curves. By a theorem of Tissot, the characteristic directions are those of the unique orthogonal net defined by

$$
(5) \qquad Fy'^2 + (E - G)y' - F = 0,
$$

on Π which by the non-conformal transformation T is converted into an orthogonal net on Σ. This is termed the characteristic net.

4. The terminal ellipse. If, on the plane Π, the scale function σ is laid off along the corresponding direction y' at a fixed point (x, y), the end point will describe a locus, which is an ellipse. Let (X, Y) represent the cartesian coordinates of a point on the ellipse relative to the fixed point (x, y), which is, of course, the center of the ellipse. Then $\sigma^2 = X^2 + Y^2$ and $y' = Y/X$. By (3), it is found that the equation of this terminal ellipse is

$$EX^2 + 2FXY + GY^2 = 1. \tag{6}$$

The directions of the major and minor axes of this terminal ellipse are given by (5) so that they coincide with the characteristic directions. The semi-major and semi-minor diameters of the terminal ellipse are solutions of the equation

$$H^2\sigma^4 - (E + G)\sigma^2 + 1 = 0. \tag{7}$$

For a fixed σ, the preceding equation defines a curve such that at all points of this curve the corresponding terminal ellipses are congruent. This set of $2\infty^1$ curves is termed the *singular family* of curves.

It is noted that this singular family of $2\infty^1$ curves can degenerate into only ∞^1 curves if and only if, either, the mapping is conformal, or else the functions $E + G$ and $H^2 = EG - F^2$, are functionally related. For example, under any area preserving map the singular family is $E + G = \text{const.}$

5. Characterization of the system of ∞^2 scale curves. Let (X, Y) represent the cartesian coordinates relative to the point (x, y) of the center of curvature constructed at the point (x, y) of a scale curve. Then $X = -y'(1 + y'^2)/y''$, $Y = (1 + y'^2)/y''$. From these we find $y' = -X/Y$, $y'' = (X^2 - Y^2)/Y^3$. Substituting these values of y' and y'' into the differential equation (4) and simplifying, we obtain the equation

$$\begin{aligned} G_yX^3 - (2F_y + G_x)X^2Y + (E_y + 2F_x)XY^2 - E_xY^3 \\ + 2[-FX^2 + (E - G)XY + FY^2] = 0. \end{aligned} \tag{8}$$

THEOREM 2. *The locus of the centers of curvature (the central locus) is a general cubic curve which has a node at the fixed point (x, y), the directions of the tangent lines at the node coinciding with the characteristic directions.*

This property is possessed not only by any family of ∞^2 scale curves but also by a more general class of families of curves. If a system of curves possesses the preceding property, then it must be defined by a differential equation of the form

$$y'' = \frac{(1 + y'^2)(\alpha + \beta y' + \gamma y'^2 + \delta y'^3)}{2(-\eta + \epsilon y' + \eta y'^2)}, \tag{9}$$

where $(\alpha, \beta, \gamma, \delta, \epsilon, \eta)$ are six functions of (x, y).

THEOREM 3. *The differential equation* (9) *will represent the* ∞^2 *scale curves of a non-conformal map of a surface* Σ *upon a plane* Π *if, and only if the six functions* $(\alpha, \beta, \gamma, \delta, \epsilon, \eta)$ *of* (x, y) *satisfy the two partial differential equations of the second order*

$$(10) \qquad \begin{aligned} &M_y = N_x, \\ &\alpha_y + N\alpha = \delta_x + \epsilon_{xy} + N\epsilon_x + \epsilon N_x + M(\delta + \epsilon_y + \epsilon N), \end{aligned}$$

where M *and* N *are defined by*

$$(11) \qquad \begin{aligned} (\epsilon^2 + 4\eta^2)M &= \epsilon(\alpha - \gamma) + 2\eta(\beta - \delta) - 4\eta\eta_x + 2\epsilon\eta_y - \epsilon\epsilon_x - 2\eta\epsilon_y, \\ (\epsilon^2 + 4\eta^2)N &= -2\eta(\alpha - \gamma) + \epsilon(\beta - \delta) - 2\epsilon\eta_x - 4\eta\eta_y + 2\eta\epsilon_x - \epsilon\epsilon_y. \end{aligned}$$

To prove this proposition, we proceed as follows. Upon comparing (9) with the differential equation (4) of the ∞^2 scale curves of a non-conformal cartogram, it follows that there exists a function $\rho(x, y) \neq 0$, such that

$$(12) \qquad \begin{gathered} F = \rho\eta, \quad E - G = \rho\epsilon, \quad E_x = \rho\alpha, \quad E_y + 2F_x = \rho\beta, \\ 2F_y + G_x = \rho\gamma, \quad G_y = \rho\delta. \end{gathered}$$

From these equations we eliminate F and G and obtain

$$(13) \qquad \begin{aligned} E_x &= \rho\alpha = \rho(\gamma - 2\eta_y + \epsilon_x) + \epsilon\rho_x - 2\eta\rho_y, \\ E_y &= \rho(\beta - 2\eta_x) - 2\eta\rho_x = \rho(\delta + \epsilon_y) + \epsilon\rho_y. \end{aligned}$$

These equations can be solved for ρ_x/ρ and ρ_y/ρ, giving the solutions

$$(14) \qquad \rho_x/\rho = M, \qquad \rho_y/\rho = N,$$

where M and N are defined by (11).

Substituting these into (13), we find

$$(15) \qquad \begin{aligned} E_x &= \rho\alpha = \rho(\gamma - 2\eta_y + \epsilon_x + \epsilon M - 2\eta N), \\ E_y &= \rho(\beta - 2\eta_x - 2\eta M) = \rho(\delta + \epsilon_y + \epsilon N). \end{aligned}$$

Upon imposing the compatibility conditions on equations (14) for ρ and on equations (15) for E, we obtain the conditions (10). This completes the proof of our Theorem 3.

COROLLARY. *There are essentially* ∞^2 *surfaces which possess the same family of* ∞^2 *scale curves.*

A consideration of equations (12), (14), and (15) will show that if the surface Σ with the linear element (1) is mapped upon the plane Π such that

the differential equation (4) represents the scale curves, then the linear element dS' of any other surface Σ' with the same scale curves must be given by

$$dS'^2 = a dS^2 + b ds^2, \tag{16}$$

where $a \neq 0$ and b are constants. This proves the corollary stated above.

THEOREM 4. *The only point transformations which convert the class of systems of scale curves (or the class (9)) into itself, are those of the conformal group.*

This result follows from the fact that if the classes (4) or (9) are preserved, then certainly the minimal lines are preserved. That systems of scale curves are converted under conformalities into systems of scale curves, follows from a consideration of the scale function σ, given by (3).

It is remarked that the results of this section are valid only for the general case where the fraction of (4) is not reducible to lower terms.

6. Scale curves of the velocity type. Before continuing further, we shall consider briefly the second order differential equations of the cubic type and also of the velocity type.

A differential equation of the cubic type is of the form $y'' = \alpha + \beta y' + \gamma y'^2 + \delta y'^3$, where $(\alpha, \beta, \gamma, \delta)$ are functions of (x, y) only. This has been studied extensively by Lie, R. Liouville, Tresse, Kasner, and Wilczynski. With respect to the group of arbitrary point transformations, this type is called a differential equation of the first rank.

A general velocity system is defined by the differential equation $y'' = (1 + y'^2)(\psi - y'\phi)$, where ϕ and ψ are functions of (x, y). This has the property that the central locus is always a straight line.

THEOREM 5. *A system of ∞^2 scale curves can be of the cubic (Lie-Liouville) type if and only if it is a velocity system. The necessary and sufficient conditions for this are*

$$\begin{aligned} G_y(E - G) &= F(E_x + G_x + 2F_y), \\ -E_x(E - G) &= F(E_y + G_y + 2F_x). \end{aligned} \tag{17}$$

From (4), it is evident that $(1 + y'^2)$ and the denominator can not have a common root in y', since the roots of the denominator are both real. Hence for (4) to be of the cubic type, the denominator must be a factor of the cubic factor of the numerator and this demonstrates that it must be a velocity system.

Hence (4) must be of the form $y'' = (1 + y'^2)(\psi - y'\phi)/2$, and we find

$$(18) \qquad \begin{aligned} \psi F &= G_y, & \psi(E - G) &= E_x + G_x + 2F_y, \\ \phi F &= E_x, & -\phi(E - G) &= E_y + G_y + 2F_x. \end{aligned}$$

The elimination of ϕ and ψ from these equations leads to the conditions (17).

Note that if the parametric curves of the surface Σ are orthogonal, that is, $F = 0$, and if it be demanded that y'' be a polynomial in y', it does not follow necessarily that the scale curves must be of the cubic type and hence a velocity system. In this case, if we impose the condition that it be of the cubic type, then $G_y = E_x = 0$, and hence it is a velocity system. However, if the parametric curves of the surface Σ are not orthogonal, that is, $F \neq 0$, and if it be required that y'' be a polynomial in y', it does follow that the ∞^2 scale curves are of the cubic type.

Under the conditions of Theorem 5, if $F \neq 0$, the differential equation of the ∞^2 scale curves is

$$(19) \qquad y'' = (1/2F)(1 + y'^2)(-E_x + y'G_y);$$

and if $F = 0$, we find

$$(20) \qquad y'' = [1/2(E - G)](1 + y'^2)(E_y + y'G_x).$$

THEOREM 6. *The conditions* (17) *are the necessary and sufficient conditions for the cubical locus* (8) *to degenerate into three straight lines, two of which must be the tangent lines of the characteristic curves at the fixed point.*

This may be obtained by imposing the condition on (8) that it consist of three straight lines. If $F \neq 0$, the third straight line is

$$(21) \qquad G_yX + E_xY - 2F = 0;$$

and, if $F = 0$, the third straight line is

$$(22) \qquad -G_xX + E_yY + 2(E - G) = 0.$$

THEOREM 7. *If the scale curves form a natural family, then the characteristic curves are an isothermal net.*

To prove this result, we use (9) and Theorem 3. For (9) to represent the velocity system $y'' = (1 + y'^2)(\psi - y'\phi)/2$, we find

$$(23) \qquad \alpha = -\psi\eta, \quad \beta = \psi\epsilon + \phi\eta, \quad \gamma = \psi\eta - \epsilon\phi, \quad \delta = -\phi\eta.$$

Substituting these values in (11), we find that M and N are given by

$$(24) \qquad \begin{aligned} M &= \phi - \tfrac{1}{2}(\partial/\partial x)\log(\epsilon^2 + 4\eta^2) + (\partial/\partial y)\arctan(2\eta/\epsilon), \\ N &= \psi - \tfrac{1}{2}(\partial/\partial y)\log(\epsilon^2 + 4\eta^2) - (\partial/\partial x)\arctan(2\eta/\epsilon). \end{aligned}$$

Placing these values into the first of conditions (10), we have

$$(25) \qquad -\phi_y + \psi_x = (\partial^2/\partial x^2 + \partial^2/\partial y^2) \operatorname{arc\,tan} (2\eta/\epsilon).$$

The condition for a natural family is $\phi_y = \psi_x$. From this equation, it follows, by a theorem of Lie concerning isothermal families, that the characteristic curves form an isothermal net.

THEOREM 8. *If the scale curves of a non-conformal cartogram form a velocity system and if the characteristic curves are an isothermal net, then the scale curves are a natural family.*

For if the characteristic curves are an isothermal family then the right-hand side of (25) is zero. Hence $\phi_y = \psi_x$ and the scale curves form a natural family.

7. Non-conformal cartograms with rectilinear scales. We shall state and prove the following result. We obtain a new class of surfaces.

THEOREM 9. *The class of surfaces Σ for which there exists a map upon the plane Π such that the scale curves coincide with the totality of ∞^2 straight lines is*

$$(26) \quad dS^2 = a_0(ydx - xdy)^2 + 2(a_1dx + b_1dy)(ydx - xdy) + a_2dx^2 + 2c_2dxdy + b_2dy^2.$$

Thus there are essentially ∞^6 such non-conformal cartograms.

For this class of surfaces, we find

$$(27) \quad H^2 = (a_0a_2 - a_1^2)x^2 + 2(a_0c_2 - a_1b_1)xy + (a_0b_2 - b_1^2)y^2 + 2(a_1c_2 - a_2b_1)x + 2(a_1b_2 - b_1c_2)y + (a_2b_2 - c_2^2).$$

By (2), the gaussian curvature G of this class of surfaces is

$$(28) \quad H^4G = -2a_0H^2 + a_0(c_2^2 - a_2b_2) + a_2b_1^2 - 2a_1b_1c_2 + a_1^2b_2.$$

The proof of Theorem 9 is as follows. If the scale curves, represented by the differential equation (4), are all straight lines, we must have

$$(29) \quad E_x = 0, \quad E_y + 2F_x = 0, \quad 2F_y + G_x = 0, \quad G_y = 0.$$

The condition of compatibility for F yields $-2F_{xy} = E_{yy} = G_{xx}$. These together with other conditions yield

$$(30) \quad E = a_0y^2 + 2a_1y + a_2, \quad G = a_0x^2 - 2b_1x + b_2, \quad F = -a_0xy - a_1x + b_1y + c_2.$$

These then lead to the equation (26) and the proof is complete.

THEOREM 10. *The only surfaces* Σ *of constant gaussian curvature of the class* (26) *are the developable ones. Thus there exists no non-conformal map of a sphere or pseudo-sphere upon a plane such that the scale curves coincide with the totality of the* ∞^2 *straight lines.*

By (28), it follows that if G is identically constant, then H is identically constant. By (27), we must have

$$(31)\quad a_0a_2 - a_1^2 = a_0c_2 - a_1b_1 = a_0b_2 - b_1^2 = a_1c_2 - a_2b_1 = a_1b_2 - b_1c_2 = 0.$$

If $a_0 \neq 0$, these five equations will be satisfied by

$$(32)\qquad a_2 = a_1^2/a_0, \quad c_2 = a_1b_1/a_0, \quad b_2 = b_1^2/a_0.$$

However in this case, we shall have $H^2 = a_2b_2 - c_2^2 = 0$. This contradiction shows that $a_0 = 0$.

Since $a_0 = 0$, it follows by (31) that $a_0 = a_1 = b_1 = 0$. By (28), it is seen that G is identically zero. Hence the surface Σ is developable and Theorem 10 is proved.

It is remarked that for a developable surface Σ of Theorem 10, the map T is an unrolling of Σ upon the plane Π followed by an affine transformation in Π. To a given scale σ, corresponds a parallel pencil of straight lines.

Now let us consider the general class of surfaces (26) *for which* $a_0 \neq 0$.

By an appropriate similitude, the class (26) may be reduced to the form

$$(33)\qquad dS^2 = (ydx - xdy)^2 + adx^2 + bdy^2.$$

By (27) and (28), we find

$$(34)\qquad H^2 = ax^2 + by^2 + ab, \qquad H^4G = -2H^2 - ab.$$

It is observed that a and b cannot both be zero. If one of these is zero, then the other must be positive. The locus $H = 0$ can be an imaginary or real ellipse according as a and b are both positive or both negative, a hyperbola if a and b are of oppositive signs, and two coincident straight lines if a or b is zero.

The curves $G = \text{const.}$, are conic sections similar and similarly placed to $H = 0$.

The surface Σ exists for all values of (x, y) if $H = 0$ is an imaginary ellipse, for the interior points of the real ellipse $H = 0$, for the points of the side of the hyperbola $H = 0$ not containing the center, and for all the points on both sides of the line $H = 0$.

The scale σ satisfies the equation

$$(y - xy')^2 + a + by'^2 = (1/\sigma^2)(1 + y'^2). \tag{35}$$

Hence to a given scale σ correspond ∞^1 straight lines. These straight lines will all be tangent to the conic

$$x^2/(1/\sigma^2 - b) + y^2/(1/\sigma^2 - a) = 1, \tag{36}$$

confocal with the conic $H = 0$.

Finally let us consider the general class of surfaces (26) *for which* $a_0 = 0$.

By an appropriate similitude, the class (26) may be reduced to the form

$$dS^2 = 2dy(ydx - xdy) + adx^2. \tag{37}$$

By (27) and (28), we find

$$H^2 = -y^2 - 2ax, \qquad H^4G = a. \tag{38}$$

It is noted that $H = 0$ is a parabola. The surface Σ exists for all points in the interior of this parabola $H = 0$.

The curves $G =$ const. are parabolas congruent to the parabola $H = 0$ and having the same axis as that of $H = 0$.

The scale σ satisfies the equation

$$2y'(y - xy') + a = (1/\sigma^2)(1 + y'^2). \tag{39}$$

Hence to a given scale σ correspond ∞^1 straight lines. These straight lines will be all tangent to the parabola

$$y^2 + (a - 1/\sigma^2)(2x + 1/\sigma^2) = 0, \tag{40}$$

confocal with the parabola $H = 0$.

THEOREM 11. *In the general class* (26), *where at least one of the constants* a_0, a_1, b_1, *is not zero, the surface* Σ *exists on one side of the conic section* $H = 0$. *The gaussian curvature* G *is constant along the conic sections similar and similarly placed to the ellipse or hyperbola* $H = 0$, *or along the parabolas congruent to and having the same axis as* $H = 0$, *or along the lines parallel to the line* $H = 0$. *To a given scale* σ *there correspond* ∞^1 *straight scales all tangent to a conic confocal with* $H = 0$.

This theorem summarizes all the work of the last few paragraphs.

Our new general theory which has been developed in this paper will be applied to the classic map projections or cartograms (area-preserving, azimuthal, etc.) in later papers. The family of ∞^2 scale curves is usually quite complicated.

COLUMBIA UNIVERSITY,
ILLINOIS INSTITUTE OF TECHNOLOGY.

BIBLIOGRAPHY.

1. E. Kasner, "Differential geometric aspects of dynamics," *American Mathematical Society Colloquium Publications*, vol. 3 (1912, 1934); *Transactions of the American Mathematical Society*, vol. 9, vol. 10 (1908-1909).

2. E. Kasner, "The geometry of differential elements of the second order with respect to the group of all point transformations," *American Journal of Mathematics*, vol. 28 (1907), pp. 203-213.

3. E. Kasner and J. De Cicco, "The geometry of velocity systems," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 236-245.

4. J. De Cicco, "New proofs of the theorems of Beltrami and Kasner on linear families," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 407-412.

5. E. Kasner, "Geometric properties of isothermal families," *Institute de Matematica*, Rosario, Argentina, vol. 5 (1943).

6. E. Kasner and J. De Cicco, "Ovals should be used to map airplane ranges," *Science News Letter*, March 25, 1944, p. 200.

7. E. Kasner and J. De Cicco, "Scale curves in conformal maps," *Proceedings of the National Academy of Sciences*, vol. 30 (1944), pp. 162-164.

8. E. Kasner and J. De Cicco, "Scale curves in general cartography," *Proceedings of the National Academy of Sciences*, vol. 30 (1944), pp. 211-215.

9. E. Kasner and J. De Cicco, "Geometry of scale curves in conformal maps," *American Journal of Mathematics*, vol. 67 (1945), pp. 157-166.

10. J. De Cicco, "Conformal maps of surfaces with isothermal systems of scale curves," *American Journal of Mathematics*, vol. 68 (1946), pp. 137-146.

PROJECTIVE ALGEBRA I.*

C. J. Everett and S. Ulam.

1. Projective algebra of subsets of a direct product. Let X and Y be any two sets of points and $[X;Y]$ the class of all pairs $[x;y]$, $x \epsilon X$, $y \epsilon Y$. Fix $[x_0;y_0]$ arbitrarily in $[X;Y]$. Suppose now that $\mathfrak{B}$ is a boolean algebra of subsets A of $[X;Y]$, including as the identity I the entire set $[X;Y]$, and the null set 0. Define for every $A \epsilon \mathfrak{B}$, the set A_x as the class of all pairs $[x;y_0]$ for which there exists a $y \epsilon Y$ such that $[x;y] \epsilon A$. Define A_y similarly. For every $A = [X_1;y_0]$ in $\mathfrak{B}$, $X_1 \subset X$, and $B = [x_0;Y_1] \epsilon \mathfrak{B}$, $Y_1 \subset Y$, define the "product" set $A \square B$ as the class of all pairs $[x_1;y_1]$, $x_1 \epsilon X_1$, $y_1 \epsilon Y_1$. Note that if A or B is 0, $A \square B = 0$. We say that $\mathfrak{B}$ is a *projective algebra of subsets of the product* $[X;Y]$ in case for all A of $\mathfrak{B}$, A_x and A_y also are in $\mathfrak{B}$, and $A \square B$ is in $\mathfrak{B}$ for A, B of the above type. Thus we demand that the boolean algebra $\mathfrak{B}$ be closed under projection and $\square$-product formation. In particular, the boolean algebra of *all* subsets of $[X;Y]$ is such a projective algebra.

It may be remarked that the following is a modest beginning for a study of logic with quantifiers from a boolean point of view, since, for example, the set A_x is essentially (the y_0 being a dummy) the class of all x for which there exists a y such that the proposition $A(x,y)$ is true.

Our object is to discover a set of properties of the above model which, adopted as postulates for an abstract "projective algebra," will permit a representation theorem to the effect that every such abstract algebra is isomorphic, with preservation of union, intersection, complement, x- and y-projection, and $\square$-product, to a projective algebra of subsets of some direct product $[X;Y]$. This is accomplished only for the atomic case in the present paper. The authors hope to study the general problem and related matters subsequently.

We have derived many of the more immediate properties of the abstractly defined projective algebra, and have shown, in particular, that every such algebra is embeddable in a complete ordered projective algebra.

The essential properties of the above model which we use later as axioms are the following.

Property 1. $(A \cup B)_x = A_x \cup B_x$; $(A \cup B)_y = A_y \cup B_y$.

* Received February 16, 1945.

For if $[x; y_0] \in (A \cup B)_x$, then for some y, $[x; y]$ is in A or B; hence $[x; y_0]$ is in A_x or B_x. Similarly, $A_x \cup B_x \subset (A \cup B)_x$. If either A or B is 0, Property 1 is trivial. (See Property 3.)

PROPERTY 2. *$I_{xy} = [x_0; y_0] = I_{yx}$, where $[x_0; y_0]$ is an atom in $\mathfrak{B}$, i. e., an element minimal over 0 in $\mathfrak{B}$.*

For $I_x = [X; y_0]$, and I_{xy} is the class of all $[x_0; y]$ such that for some x, $[x; y] \in I_x$; hence $I_{xy} = [x_0; y_0]$.

PROPERTY 3. *$A_x = 0$ if and only if $A = 0$, and similarly for y-projection.*

If $A = 0$, then the set of all $[x; y_0]$ for which there exists an $[x; y]$ in A is empty. If $A \neq 0$, there is at least one point $[x; y] \in A$; and, hence, at least one $[x; y_0]$ in A_x.

PROPERTY 4. *$A_{xx} = A_x$; $A_{yy} = A_y$.*

If $[x_1; y_0] \in A_x$, then $[x_1, y_0] \in A_{xx}$. If $[x_1; y_0] \in A_{xx}$ then for some y, $[x_1; y] \in A_x$; hence $y = y_0$, and $[x_1; y_0] \in A_x$. If $A = 0$, Property 4 is trivial.

PROPERTY 5. *For $0 \neq A = [X_1; y_0] \subsetneq I_x$, $0 \neq B = [x_0; Y_1] \subset I_y$, the set $A \square B = \{[x_1; y_1]; x_1 \in X_1, y_1 \in Y_1\}$ has the property*

$$(A \square B)_x = A, \qquad (A \square B)_y = B,$$

and if $S \in \mathfrak{B}$, $S_x = A$, $S_y = B$, then $S \subset A \square B$.

For $(A \square B)_x$ is the set of all $[x; y_0]$ such that $[x; y] \in [X_1; Y_1]$ for some y; hence $(A \square B)_x = [X_1; y_0] = A$. Suppose now $S_x = [X_1; y_0]$, $S_y = [x_0; Y_1]$ and $[x; y] \in S$. Then $[x; y_0] \in S_x = [X_1; y_0]$ and $[x_0; y] \in S_y = [x_0; Y_1]$. Hence $x \in X_1$, $y \in Y_1$, and $[x; y] \in [X_1; Y_1] = A \square B$. Thus $S \subset A \square B$.

PROPERTY 6. *$I_x \square p_0 = I_x$, and $p_0 \square I_y = I_y$, where $p_0 = [x_0; y_0]$.*

This is immediate from the $\square$-definition.

PROPERTY 7. *For $A_1 = [X_1; y_0]$, $A_2 = [X_2; y_0]$ in $\mathfrak{B}$, one has $(A_1 \cup A_2) \square I_y = (A_1 \square I_y) \cup (A_2 \square I_y)$, and similarly for $B_i = [x_0; Y_i]$ and I_x.*

For $(A_1 \cup A_2) \square I_y = [X_1 \cup X_2; y_0] \square [x_0; Y] = [X_1 \cup X_2; Y]$ and $(A_1 \square I_y) \cup (A_2 \square I_y) = [X_1; Y] \cup [X_2; Y] = [X_1 \cup X_2; Y]$.

If A_1 or A_2 is 0, Property 7 is trivial.

2. Projective algebra. Let $\mathfrak{B}$ be a boolean algebra with unit i, zero 0, $i > 0$, so that for all $a \in \mathfrak{B}$, $0 \leqq a \leqq i$. $\mathfrak{B}$ is said to be a *projective algebra*

if two mappings $a \to a_x$ and $a \to a_y$ of $\mathfrak{B}$ into $\mathfrak{B}$ are defined, satisfying the following postulates.

P1. $(a \vee b)_x = a_x \vee b_x$; $(a \vee b)_y = a_y \vee b_y$.

P2. $i_{xy} = p_0 = i_{yx}$ *where* p_0 *is an atom of* $\mathfrak{B}$, *that is, an element minimal over* 0 *in* $\mathfrak{B}$.

P3. $a_x = 0$ *if and only if* $a = 0$, *and* $a_y = 0$ *if and only if* $a = 0$.

P4. $a_{xx} = a_x$; $a_{yy} = a_y$.

P5. For $0 < a \leqq i_x$, $0 < b \leqq i_y$, *there exists an element* $a \square b$ *such that* $(a \square b)_x = a$, $(a \square b)_y = b$, *with the property that* $t \in \mathfrak{B}$, $t_x = a$, $t_y = b$ *implies* $t \leqq a \square b$.

P6. $i_x \square p_0 = i_x$; $p_0 \square i_y = i_y$.

P7. $0 < a_1, a_2 \leqq i_x$ *implies* $(a_1 \vee a_2) \square i_y = (a_1 \square i_y) \vee (a_2 \square i_y)$; *and* $0 < b_1, b_2 \leqq i_y$ *implies* $i_x \square (b_1 \vee b_2) = (i_x \square b_1) \vee (i_x \square b_2)$.

We prove a number of immediate consequences of these postulates.

C1. $a \leqq b$ *implies* $a_x \leqq b_x$ and $a_y \leqq b_y$.

From $a \vee b = b$ one has (*P1*) $(a \vee b)_x = b_x = a_x \vee b_x$.

C2. For all $a \in \mathfrak{B}$, $a_x \leqq i_x$, $a_y \leqq i_y$.

By *C1*, $a \leqq i$; hence $a_x \leqq i_x$.

C3. $i_x \wedge (a_x)' \leqq (a')_x$ *and similarly for y-projection.*

For $i = a \vee a'$, $i_x = a_x \vee (a')_x$ (*P1*), $i_x \wedge (a_x)' = (a_x \vee (a')_x) \wedge (a_x)'$ $= 0 \vee ((a')_x \wedge (a_x)') \leqq (a')_x$.

C4. $(a \wedge b)_x \leqq a_x \wedge b_x$; $(a \wedge b)_y \leqq a_y \wedge b_y$.

For $(a \wedge b) \leqq a, b$, hence (*C1*) $(a \wedge b)_x \leqq a_x, b_x$.

C5. $a > 0$ *implies* $a_{xy} = p_0 = a_{yx}$.

Since $a \leqq i$, $a_{xy} \leqq i_{xy} = p_0$ (*C1*, *P2*). Since p_0 is an atom, $a_{xy} = p_0$ for $a_{xy} = 0$ is impossible by *P3*.

C6. $(p_0)_x = p_0 = (p_0)_y \leqq i_x \wedge i_y$.

For $p_0 = i_{xy}$ and hence $(p_0)_y = i_{xyy} = i_{xy} = p_0$, (*P2*, *P4*). The final inequality follows from *C2*.

In connection with the $\square$-product, it is convenient to make two definitions.

D1. For $a \leqq i_x$, $b \leqq i_y$, define $a \square 0$, $0 \square b$, and $0 \square 0$ all to be 0.

It is seen that for all of these the x- and y-projections are 0.

D2. For all $a \in \mathfrak{B}$, define the "closure" a^* of a to be $a_x \square a_y$.

C7. $a \square b = 0$ *if and only if* $a = 0$ *or* $b = 0$.

For if $a > 0$, $b > 0$, by *P5*, $(a \square b)_x = a > 0$, hence $a \square b > 0$ by *P3*.

C8. The following are equivalent: (a) $a \leqq i_x$; (b) *there is a* $t \in \mathfrak{B}$ *such that* $t_x = a$; (c) $a_x = a$; *and similarly for y-projections.*

(a) implies (b), for if $a = 0$, $a_x = a$ by $P3$, whereas for $a > 0$, $(a \square i_y)_x = a$ by $P5$. Next, (b) implies (c), since $t_x = a$ yields $t_{xx} = a_x = t_x = a$ ($P4$). Finally, (c) implies (a) by $C2$.

C9. $0 < a_1 \leqq a_2 \leqq i_x$, $0 < b_1 \leqq b_2 \leqq i_y$ *if and only if* $0 < a_1 \square b_1 \leqq a_2 \square b_2$.

If $0 < a_1 \square b_1 \leqq a_2 \square b_2$, then ($C1$, $P5$) $0 < a_1 \leqq a_2$, $0 < b_1 \leqq b_2$. Conversely, assume the first inequalities and let $u = (a_1 \square b_1) \vee (a_2 \square b_2)$. Then $u_x = a_1 \vee a_2 = a_2$, $u_y = b_1 \vee b_2 = b_2$ ($P1$, $P5$); hence $u \leqq a_2 \square b_2$, by $P5$, and $a_1 \square b_1 \leqq a_2 \square b_2$.

C10. The (*) *operation has the following closure properties*:

$$0^* = 0; \quad a^* \geqq a; \quad a^{**} = a^*; \quad a \geqq b \text{ implies } a^* \geqq b^*; \quad i^* = i.$$

First, $0^* = 0_x \square 0_y = 0 \square 0 = 0$. Second, $a^* = a_x \square a_y \geqq a$ by the maximality property of $\square$-product in $P5$, for $a > 0$. For $a = 0$, this becomes trivial. Third, $a^{**} = (a_x \square a_y)_x \square (a_x \square a_y)_y = a_x \square a_y = a^*$, for $a \geqq 0$. Fourth, $a \leqq b$ implies $a^* \leqq b^*$ by $C1$, $C9$. Fifth, $i^* \geqq i$, hence $i^* = i$.

C11. $a_1, a_2 \leqq i_x$, $b_1, b_2 \leqq i_y$ *imply* $(a_1 \wedge a_2) \square (b_1 \wedge b_2) = (a_1 \square b_1) \wedge (a_2 \square b_2)$.

If any one of the a_i, b_i is 0, $C11$ is trivial. Otherwise, $a_1 \wedge a_2 \leqq a_1, a_2$, $b_1 \wedge b_2 \leqq b_1, b_2$, hence $(a_1 \wedge a_2) \square (b_1 \wedge b_2) \leqq a_1 \square b_1$, $a_2 \square b_2$ by $C9$. Hence the left member is contained in the right. Let t denote the right member. We have $t_x \leqq a_1 \wedge a_2$, $t_y \leqq b_1 \wedge b_2$ ($C4$, $P5$); hence $t \leqq t^* = t_x \square t_y \leqq (a_1 \wedge a_2) \square (b_1 \wedge b_2)$ by $C10$, $C9$.

C12. $a \leqq i_x$, $b \leqq i_y$ *implies* $a \square b = (a \square i_y) \wedge (i_x \square b)$.

In $C11$, let $a_1 = a$, $a_2 = i_x$, $b_1 = i_y$, $b_2 = b$.

C13. $t \leqq a \square b$ *implies* $t_x \leqq a$, $t_y \leqq b$.

This requires only $C1$ and $P5$ for $t > 0$. It is trivial for $t = 0$.

C14. $((a_1 \square b_1) \vee (a_2 \square b_2))^* = (a_1 \vee a_2) \square (b_1 \vee b_2)$ *for* a_i, b_i *not* 0.

We require $P1$, $P5$.

C15. $a \leqq i_x$ *implies* $(a \square i_y) \wedge i_x = a$.

For $a \leqq i_x$, $a \leqq a^* = a \square p_0 < a \square i_y$. Hence $a \leqq (a \square i_y) \wedge i_x$. Now let d be this intersection. Then $d \leqq i_x$, $d_x = d \leqq a \wedge i_x = a$.

C16. All a, b *satisfying* $0 < a \leqq i_x$, $0 < b \leqq i_y$ *are closed, i. e.*, $a \square p_0 = a$, $p_0 \square b = b$.

$a^* = a \square p_0$ since $a_x = a$ and $a_y = a_{xy} = p_0$. By $C12$, $a \square p_0 = (a \square i_y) \wedge (i_x \square p_0) = (a \square i_y) \wedge i_x$ by $P6$. Then $a \square p_0 = a$ by $C15$.

C17. $(p_0)^* = p_0 = p_0 \square p_0 = (p_0 \square i_y) \wedge (i_x \square p_0) = i_x \wedge i_y$.

This follows from $C6$, $C16$, $C12$, $P6$.

$C18$. *If $t_y = p_0$ then $t \leqq i_x$. If $t_x = p_0$ then $t \leqq i_y$.*

For $t \leqq t^* \leqq i_x \square p_0 = i_x$ by $C10$, $C9$, $P6$.

$C19$. *For $a_1, a_2 \leqq i_x$, $b \leqq i_y$, one has $(a_1 \vee a_2) \square b = (a_1 \square b) \vee (a_2 \square b)$. Similarly for $b_1, b_2 \leqq i_y$, $a \leqq i_x$.*

By $C12$, $P7$, $(a_1 \vee a_2) \square b = ((a_1 \vee a_2) \square i_y) \wedge (i_x \square b) = ((a_1 \square i_y) \vee (a_2 \square i_y)) \wedge (i_x \square b) = ((a_1 \square i_y) \wedge (i_x \square b)) \vee ((a_2 \square i_y) \wedge (i_x \square b)) = (a_1 \square b) \vee (a_2 \square b)$.

$C20$. *For $a_1, a_2 \leqq i_x$, $b_1, b_2 \leqq i_y$ one has $(a_1 \vee a_2) \square (b_1 \vee b_2) = (a_1 \square b_1) \vee (a_1 \square b_2) \vee (a_2 \square b_1) \vee (a_2 \square b_2)$.*

This requires two applications of $C19$.

$C21$. $(a \square i_y)' = (a' \wedge i_x) \square i_y$; $(i_x \square b)' = i_x \square (b' \wedge i_y)$.

For $(a \square i_y) \vee ((a' \wedge i_x) \square i_y) = (a \vee (a' \wedge i_x)) \square i_y = i_x \square i_y = i$, by $P7$, and $(a \square i_y) \wedge ((a' \wedge i_x) \square i_y) = (a \wedge (a' \wedge i_x)) \square i_y = 0 \square i_y = 0$, by $C11$. Since complementation is unique in a boolean algebra, $C21$ follows.

$C22$. $(a \square b)' = ((a' \wedge i_x) \square i_y) \vee (i_x \square (b' \wedge i_y))$.

This follows from $C12$ and $C21$.

$C23$. *For $c = c^*$, $c_y < i_y$, one has $(c')_x = i_x$.*

If $c = 0$, this is trivial. If $0 < c = c_x \square c_y$, then $c' = (((c_x)' \wedge i_x) \square i_y) \vee (i_x \square ((c_y)' \wedge i_y))$ by $C22$, and $(c')_x = (((c_x)' \wedge i_x) \square i_y)_x \vee i_x = i_x$ by $P1$, $P5$. We need to know that $(c_y)' \wedge i_y \neq 0$ but this is so since $c_y < i_y$.

$C24$. *$a \leqq i_x$, $p_0 < i_y$ implies $(a')_x = i_x$.*

For by $C16$, $a = a^*$, and $a_x = a$, $a_{xy} = p_0 = a_y$. $C24$ now follows from $C23$. (For $a = 0$, $C24$ is trivial).

$C25$. *If $p_0 < i_y$ then $((i_x)')_x = i_x$, and similarly for i_y.*

Let $a = i_x$ in $C24$.

$C26$. *$p_0 = i_y$ if and only if $i = i_x$.*

If $p_0 = i_y$, then $i = i_x \square i_y = i_x \square p_0 = i_x$ ($C10$, $P6$). If $i = i_x$, then $i_y = i_{xy} = p_0$.

$C27$. $((i_x)')_y = p'_0 \wedge i_y$.

If $p_0 = i_y$, by $C26$, $(i_x)')_y = ((i)')_y = 0_y = 0 = (i_y)' \wedge i_y = p'_0 \wedge i_y$. If $p_0 < i_y$, one has from $i_x = i_x \square p_0$ ($P6$) that $(i_x)' = (i_x \square p_0)' = i_x \square (p'_0 \wedge i_y)$ where $p'_0 \wedge i_y \neq 0$ ($C21$). Hence $((i_x)')_y = p'_0 \wedge i_y$ by $P5$.

$C28$. *The class of all x-projections is an ideal in $\mathfrak{B}$ and the correspondence $a_x \rightarrow a_x \square i_y$ is an isomorphism.*

The set of all a_x, $a \in \mathfrak{B}$ is exactly the set of all $a \leqq i_x$, by $C8$, and hence is an ideal, since the union of any two $a_1, a_2 \leqq i_x$ is $\leqq i_x$, as is $a_1 \wedge c$ for all

$c \in \mathfrak{B}$. Moreover the correspondence above preserves union ($P7$), intersection ($C11$), and complement ($C21$).

$C29$. *For* $c \in \mathfrak{B}$, $0 < a \leqq c_x$ *implies* $(a \square i_y) \wedge c > 0$.

Suppose $(a \square i_y) \wedge c = 0$. Then $c \leqq (a \square i_y)' = (a' \wedge i_x) \square i_y$. If now $a' \wedge i_x = 0$, $a = i_x$ and $(a \square i_y) \wedge c = i \wedge c = c = 0$. But then $c_x = 0$ and $a = 0$, which is a contradiction. If $a' \wedge i_x > 0$ then $a \leqq c_x \leqq a' \wedge i_x \leqq a'$ whence $a = 0$, also a contradiction.

$C30$. *If p is an atom in $\mathfrak{B}$, so are p_x and p_y.*

Suppose $0 < a < p_x$. By $C29$, $0 < (a \square i_y) \wedge p \leqq p$. Hence $(a \square i_y) \wedge p = p$ and $p \leqq a \square i_y$, $p_x \leqq a$, a contradiction.

$C31$. If $0 < c$, and $0 < p \leqq c_x$, where p is a point, then $p = ((p \square i_y) \wedge c)_x$. Moreover, if q is a point in $(p \square i_y) \wedge c$, then $q_x = p$.

This follows from $C29$, $P3$, and $C4$, thus: $0 < ((p \square i_y) \wedge c)_x \leqq p \wedge c_x = p$.

3. Completion of projective algebras.

We shall prove the following general theorem:

THEOREM 1. *If $\mathfrak{B}$ is a projective algebra, then $\mathfrak{B}$ is embeddable with preservation of (unrestricted) union, intersection, complement, x- and y-projections, and $\square$-products, in a projective algebra $\mathfrak{C}$ which is complete-ordered in the sense that every set of elements of $\mathfrak{C}$ has a l. u. b. and a g. l. b.*

Let $\mathfrak{B}$ be a projective algebra with elements $0 \leqq a \leqq i$. We recall that if $U(A)$, $L(A)$ are the sets of *all* upper and *all* lower bounds, respectively, of *all* elements of $A \subset \mathfrak{B}$, then $LU(A)$ is a closure operation.[1] Specifically, $A \subset LU(A)$, $LULU(A) = LU(A)$, (indeed one has $LUL(A) = L(A)$ and $ULU(A) = U(A)$), and $A \subset B$ implies $LU(A) \subset LU(B)$. The class $\mathfrak{C}$ of all $LU(A)$, $A \subset \mathfrak{B}$, is a complete ordered lattice which embeds $\mathfrak{B}$ under the correspondence $a \rightarrow LU(a) = L(a)$, $a \in \mathfrak{B}$, with preservation of unrestricted union and intersection. For any class of elements $LU(A_a)$ of $\mathfrak{C}$, set intersection $\cap LU(A_a)$ is effective as g. l. b. $\wedge LU(A_a)$, and *closure* of set union $LU(\cup LU(A_a))$ is effective as l. u. b. $\vee LU(A_a)$. Note that $LU(i) = L(i) = \mathfrak{B}$ contains all $LU(A) \in \mathfrak{C}$ and is thus the identity of $\mathfrak{C}$. Also $LU(0) = L(0) = (0)$ is the zero of $\mathfrak{C}$. It is emphasized that while the elements of $\mathfrak{C}$ are sets, and the order is that of set inclusion, the boolean algebra $\mathfrak{C}$ is not the ordinary one of subsets of a set, in particular, union is *not* set union, and the zero of $\mathfrak{C}$ is *not* the null set.

---

[1] G. Birkhoff, "Lattice theory," *American Mathematical Society Colloquium Publications*, vol. 25 (1940), p. 25; H. MacNeille, "Partially ordered sets," *Transactions of the American Mathematical Society*, vol. 42 (1937), pp. 416-460.

MacNeille[2] has shown that in $\mathfrak{C}$, $LU(A) \vee L(\text{all } a', a \in A) = LU(i)$, and $LU(A) \wedge L(\text{all } a') = LU(0)$, and that the correspondence $LU(A) \to L(\text{all } a')$ is a dual isomorphism of $\mathfrak{C}$ onto all of itself. From this it follows that $\mathfrak{C}$ is distributive and hence a boolean algebra, with $L(\text{all } a')$ effective as complement of $LU(A)$. The correspondence $a \to LU(a)$ obviously preserves complements, since $(LU(a))' = LU(a')$.

In this section we shall understand that if $A \subset \mathfrak{B}$, A_x is to mean the set of all a_x, $a \in A$. In $\mathfrak{C}$ define $(LU(A))_x = LU(A_x)$. We note that projection is well defined, i. e., $LU(A) = LU(B)$ implies $LU(A_x) = LU(B_x)$. It is sufficient to prove $A_x \subset LU(B_x)$. If $B = (0)$ we have $A \subset LU(A) = LU(0) = L(0) = (0)$; hence $A = (0)$, and $A_x = (0_x) = (0) \subset LU(B_x)$. Now let B contain at least one $b > 0$, and let $u \geqq B_x$. In particular, $u \geqq b_x > 0$, hence $u \wedge i_x \geqq b_x > 0$. Then $(u \wedge i_x) \square i_y) \geqq b_x \square b_y = b^* \geqq b$, all $b \in B$. But $U(A) = U(B)$, hence $(u \wedge i_x) \square i_y \geqq A$, and $u \geqq u \wedge i_x \geqq A_x$. Thus $A_x \subset LU(B_x)$. Similarly $B_x \subset LU(A_x)$. Moreover the correspondence $a \to LU(a)$ preserves projection, since $LU(a_x) = (LU(a))_x$.

Finally, in $\mathfrak{C}$ we define, for $(0) \neq LU(A) \subset LU(i_x)$, $(0) \neq LU(B) \subset LU(i_y)$, the direct product $LU(A) \square LU(B)$ as the (closed) set:

$$L(\text{all } u \square v;\ u \leqq i_x,\ v \leqq i_y,\ u \square v \geqq \text{all } \dot{a} \square \dot{b},\ \dot{a} \in LU(A),\ \dot{b} \in LU(B)).$$

The form of this definition is convenient in that it is clearly an element of $\mathfrak{C}$, since for any X, $L(X) = LUL(X)$; it is cumbersome, however, and we prove the following equivalences.

LEMMA 1. *For $LU(A) \subset LU(i_x)$, $LU(A) = L_xU_x(A)$ where the sub-x means that the operator is restricted to the elements $\leqq i_x$, that is to the elements of the sub-boolean algebra of all x-projections.*

Clearly $U(A) \supset U_x(A)$; hence $LU(A) \subset LU_x(A)$. But $LU_x(A) = L_xU_x(A)$. For $LU_x(A) \supset L_xU_x(A)$, and if $l \leqq U_x(A)$, then $l \leqq i_x$ since $i_x \geqq (A)$; (recall that $LU(i_x) = L(i_x) \supset LU(A) \supset A$). Hence $LU(A) \subset L_xU_x(A)$. Now let $l_x \leqq U_x(A)$, $u \geqq A$; then $i_x, u \geqq u \wedge i_x \geqq A$, hence $l_x \leqq u \wedge i_x \leqq u$.

LEMMA 2. $LU(A) \square LU(B) = L(\text{all } u \square v;\ u \in U_x(A),\ v \in (U_y(B)).$

For in the $\square$-definition u ranges over all $U_xL_xU_x(A) = U_x(A)$.

LEMMA 3. $LU(A) \square LU(B) = (\text{all } l;\ l \leqq \dot{a} \square \dot{b} \text{ for some } \dot{a} \in LU(A), \dot{b} \in LU(B))$.

For let $l \in L(u \square v;\ u \in U_x(A), v \in U_y(B))$. Then $l \leqq u \square v$, $l_x \leqq$ all u,

[2] H. MacNeille, *loc. cit.*

$l_y \leqq$ all v, $l_x \epsilon L_x U_x(A)$, $l_y \epsilon L_y U_y(B)$; hence $l_x = \dot{a}$, $l_y = \dot{b}$, a $= \dot{a} \square \dot{b}$. Conversely, let $l \leqq \dot{a} \square \dot{b}$, for $\dot{a} \epsilon LU(A) = L_x U_x(A)$ $= L_y U_y(B)$. Since $U_x(A) = U_x L_x U_x(A)$, $U_y L_y U_y(B)$, $\leqq$ all $u \square v$. Thus $l \epsilon L($all $u \square v)$.

It is now clear that $a \to L(U(a))$ preserves $\square$-product, $\square LU(b) = LU(a \square b) = L(a \square b)$. For $LU(a) \square LU(b) =$ $\dot{a} \square \dot{b}$, $\dot{a} \epsilon LU(a)$, $\dot{b} \epsilon LU(b)) = (l; l \leqq$ some $\dot{a} \square \dot{b}$, $\dot{a} \leqq a$, $\dot{b} \leqq$

We proceed now to verify $P1$-7 in $\mathfrak{C}$.

$P1$. $(LU(A) \vee LU(B))_x = (LU(A))_x \vee (LU(B))_x$.

If $A = (0)$ or $B = (0)$, $P1$ is trivial. Using the definition $P1$ to $LU((LU(A) \cup LU(B))_x) = LU(LU(A_x) \cup LU(B))$. (both sides with U and recalling that $ULU(A) = U(A)$, we have $\cup LU(B))_x) = U(A_x) \cap U(B_x)$. Since $A \subset LU(A)$, $B \subset LU($ B_x are in $(LU(A) \cup LU(B))_x$ and the left member is contained Now let $u \geqq A_x$, B_x, and $l \epsilon LU(A)$. Then $u \triangle i_x \geqq A$ $(u \wedge i_x) \square i_y \geqq a_x \square a_y \geqq a$ for all $a \epsilon A$. Since at least one a > 0. Thus $l \leqq (u \wedge i_x) \square i_y$ and $l_x \leqq (u \wedge i_x) \leqq u$. Hence the is in the right.

$P2$. $(LU(i))_{xy} = (LU(i))_{yx}$ is a point in $\mathfrak{C}$; indeed $LU(p_0) = (0, p_0)$.

For $(LU(i))_{xy} = LU(i_{xy}) = LU(p_0) = L(p_0) = (0, p_0$ minimal over $LU(0)$.

$P3$. $(LU(A))_x = LU(0)$ if, and only if, $LU(A) = LU(0)$

If $LU(A) = LU(0)$ clearly $(LU(A))_x = LU(0_x) = LU$ versely if $LU(A_x) = (0) \supset A_x$, then $a_x = 0$ and $a = 0$ for all $LU(A) = LU(0)$.

$P4$. $(LU(A))_{xx} = (LU(A))_x$.

The proof is trivial.

$P5$. For $(0) \neq LU(A) \subset LU(i_x)$ and $(0) \neq LU(B) \subset$ direct product has x-projection $LU(A)$, y-projection $LU(B)$, and in $\mathfrak{C}$ with these properties.

First, $LU((L(u \square v; u \square v \geqq$ all $\dot{a} \square \dot{b}$, $\dot{a} \epsilon LU(A)$, $\dot{b}$ $= LU(A)$. Let $l \leqq$ all $u \square v$, $u_0 \epsilon U(A) = ULU(A)$, hence u_0 $u_0 \triangle i_x \geqq$ all $\dot{a}$, hence $u_0 \wedge i_x > 0$ since $LU(A) \neq (0)$. Then $(u$ $\geqq$ all $\dot{a} \square \dot{b}$, and $l \leqq (u_0 \wedge i_x) \square i_y$. Thus $l_x \leqq u_0 \wedge i_x \leqq u_0$. left side is in the right. Now let $0 < a_0 \epsilon A$, $0 < b_0 \epsilon B$; the all $u \square v$, $a_0 = (a_0 \square b_0)_x$, and $A \subset (L(u \square v))_x$.

Second, suppose $(LU(S))_x = LU(A)$, $(LU(S))_y = LU(B)$, and let $s \in S$. Then $s_x \in LU(A)$, $s_y \in LU(B)$, and $u \,\square\, v \geqq s_x \,\square\, s_y \geqq s$. Thus $S \subset L(u \,\square\, v) = LU(A) \,\square\, LU(B)$.

*P*6. $LU(i_x) \,\square\, LU(p_0) = LU(i_x)$.

This follows since $\square$-product is preserved under $a \to LU(a)$.

Before proving *P*7, we note the following:

LEMMA 4. *If $S \subset LU(i_x)$, $u_1 \in LU(S)$, $u_0 \geqq S \,\square\, i_y$, then $u_1 \,\square\, i_y \leqq u_0$.*

First, if $i_x \in S$, then $u_0 \geqq i_x \,\square\, i_y = i$, and the conclusion is trivial. Suppose $i_x \notin S$. We prove $u'_0 \leqq (u_1 \,\square\, i_y)' = (u'_1 \wedge i_x) \,\square\, i_y$. Since $u'_0 \leqq (u'_0)_x \,\square\, (u'_0)_y$, it is sufficient to prove $(u'_0)_x \leqq u'_1 \wedge i_x$, i.e., $(u'_0)_x \leqq u'_1$. But it is given that $u'_0 \leqq (S' \wedge i_x) \,\square\, i_y$; hence $(u'_0)_x \leqq S' \wedge i_x \leqq S'$. For if any $s' \wedge i_x = 0$, $s = i_x \in S$ contrary to our present hypothesis. But it is given that $u_1 \leqq U(S)$; hence $u'_1 \geqq (U(S))' = L(S')$. Thus $u'_1 \geqq (u'_0)_x$.

$$P7. \quad (LU(A) \vee LU(B)) \,\square\, LU(i_y) = (LU(A) \,\square\, LU(i_y)) \vee (LU(B) \,\square\, LU(i_y)).$$

If either $A = (0)$ or $B = (0)$, *P*8 is trivially satisfied. Moreover, in any case, the right member is contained in the left by *C*9 which follows from *P*1-6.

We remark that if X and Y are any two subsets of $\mathfrak{B}$, then $LU(X \cup Y) = LU(\text{all } x \vee y;\ x \in X, y \in Y)$.

Now note the equalities: $(LU(A) \vee LU(B)) \,\square\, LU(i_y) = LU(LU(A) \cup LU(B)) \,\square\, L(i_y) = LU(\text{all } \dot a \vee \dot b;\ \dot a \in LU(A), \dot b \in LU(B)) \,\square\, L(i_y) = (\text{all } l \leqq \text{some } u \,\square\, v;\ u \in LU(\dot a \vee \dot b), v \in L(i_y)) = (\text{all } l \leqq \text{some } u \,\square\, i_y;\ u \in LU(\text{all } \dot a \vee \dot b))$.

We remark that $U(\text{all } l \leqq \text{some } \dot a \,\square\, i_y) = U(\text{all } \dot a \,\square\, i_y)$, and then note the equalities: $LU(A) \,\square\, LU(i_y) \vee (LU(B) \,\square\, LU(i_y)) = LU((\text{all } l \leqq \text{some } \dot a \,\square\, i_y) \cup (\text{all } l \leqq \text{some } \dot b \,\square\, i_y)) = L(U(l \leqq \text{some } \dot a \,\square\, i_y) \cap U(l \leqq \text{some } \dot b \,\square\, i_y)) = LU((\text{all } \dot a \,\square\, i_y) \cup (\text{all } \dot b \,\square\, i_y)) = LU(\text{all } \dot a \,\square\, i_y \vee \dot b \,\square\, i_y) = LU(\text{all } (\dot a \vee \dot b) \,\square\, i_y)$.

We must therefore prove that $(\text{all } l \leqq \text{some } u \,\square\, i_y;\ u \in LU(\text{all } \dot a \vee \dot b)) \subset LU(\text{all } (\dot a \vee \dot b) \,\square\, i_y)$. Let $l \leqq \text{some } u_1 \,\square\, i_y$, $u_1 \in LU(\text{all } \dot a \vee \dot b)$, and $u_0 \geqq \text{all } (\dot a \vee \dot b) \,\square\, i_y$. Now use Lemma 4 with $S = (\text{all } \dot a \vee \dot b)$. It follows that $u_1 \,\square\, i_y \leqq u_0$. Since $l \leqq u_1 \,\square\, i_y \leqq u_0$, the theorem is proved.

4. Representation theory. We shall prove a representation theorem for the case where $\mathfrak{B}$ is a projective algebra of *all* subsets of a set i of points p. We prove first two preliminary lemmas.

LEMMA 1. *If N is any cardinal number, there exists a (commutative) group containing exactly N elements.*

For a finite N we may use the cyclic group of order N. Now let N be infinite, and let V be a vector space of N basis elements over the galois field $GF(2)$. The elements of V may be considered as functions on a set of power N to the set $(0, 1)$ where addition is component-wise, mod 2. Denote by W the subgroup of all elements of V with only a finite number of 1-components, together with the zero element. W consists of the mutually exclusive subsets W_j of elements containing exactly j 1-components. Thus $W = \sum_0^\infty W_j$. The power of W is therefore $1 + \sum_1^\infty N^j = N$.

LEMMA 2. *If i is any (additive) group of elements p, the correspondence $p \to p + q$, q fixed, p arbitrary in i, defines a permutation π_q (one-one transformation) of i onto all of i, and the correspondence $q \to \pi_q$ is an isomorphism of i onto a group Π of permutations of the elements of i. In particular* (a) *the powers of Π and of i are equal.* (b) *for every p, $q \in i$ there exists a $\pi_s \in \Pi$ such that $\pi_s(p) = q$,* (c) *if $\pi_p(s_0) = \pi_q(s_0)$ for some $s_0 \in i$, then $\pi_p = \pi_q$.*

This is only a statement of the familiar Cayley representation of a group by permutations.

Now let i be a projective algebra of *all* subsets of a set i. We recall that $p_0 = i_x \wedge i_y$ so that all points of i are in one of the following disjoint classes:

$$\text{(a)}\ p_0, \quad \text{(b)}\ p \in i_x,\ p \neq p_0, \quad \text{(c)}\ p \in i_y,\ p \neq p_0, \quad \text{(d)}\ p \notin i_x \vee i_y.$$

Moreover it is clear (C. 30) that the points in i_x are precisely all p_x where p ranges over i; similarly for i_y. Indeed, from C. 31, it follows that a_x is exactly the set of all p_x with p in a.

Let X be the class consisting of p_0 and of all pairs (p_x, q), p_x ranging over all points $\neq p_0$ in i_x, q over all points in i. Let Y be the class consisting of p_0 and all pairs (p_y, q), p_y ranging over all points $\neq p_0$ in i_y, q over all points in i.

Define the relation $p \sim q$ on points p, q of i to mean $p_x = q_x$, $p_y = q_y$. This is an equivalence relation splitting i into mutually exclusive sets (p) of points. In each class we distinguish a special representative $\bar{p}$ (Zermelo). The points in any class $(\bar{p})$ are thus all points of i of equal closure $\bar{p}_x \square \bar{p}_y$.

We define a group on the points of i, and set up the correspondence $q \to \pi_q$ of Lemma 2, and set up the following correspondence on points of i to subsets of $[X; Y]$.

(a) $p_0 \to [p_0; p_0]$.
(b) for $p \in i_x$, $p \neq p_0$, $p \to \{[(p_x, q); p_0], q \in i\}$.
(c) for $p \in i_y$, $p \neq p_0$, $p \to \{[p_0; (p_y, q)], q \in i\}$.
(d) for $p \notin i_x \vee i_y$, and $p \in (\bar{p})$, if

$$p \neq \bar{p},\ p \to \{[(\bar{p}_x, q); (\bar{p}_y, \pi_p(q))], q \in i\}.$$

But

$$\bar{p} \to \{[(\bar{p}_x, q); (\bar{p}_y, \pi_{\bar{p}}(q))], q \in i\}$$

and

$\{[(\bar{p}_x, q); (\bar{p}_y, \pi_t(q))], q \in i, t$ an arbitrary point in the i-complement of $(\bar{p})\}$.

For any $a \leqq i$, we let $a \to A$, where A is the set of all images of all points of a under (a-d).

LEMMA 3. *If $s \to [x; y]$ and $t \to [x; y]$ under (a-d) then $s = t$.*

Different points of the type (b) map into different $[x; y]$, since the point itself appears as one of the coordinates, and similarly for type (c). If $[(\bar{p}_x, q); (\bar{p}_y, \pi_s(q))]$ is the map of two different points, they would be points with the same closure, and hence in the same $(\bar{p})$. But the pair of elements q, $\pi_s(q)$ can arise only from a unique s (Lemma 2c), and the two points would then both be p or both be $\bar{p}$, according to (d).

LEMMA 4. *Every point $[x; y]$ of $[X; Y]$ occurs as an image of some point p of i.*

For the points of $[X; Y]$ fall into the following classes:

(a) $[x; y] = [p_0; p_0]$ which is the map of p.
(b) $[x; y] = [(p_x, q); p_0]$ which is the map of $p_x \leqq i_x$, $p_x \neq p_0$.
(c) $[x; y] = [p_0; (p_y, q)]$ which is the map of $p_y \leqq i_y$, $p_y \neq p_0$.
(d) $[x; y] = [(p_x, q); (r_y, s)]$ $p_x \neq p_0$, $r_y \neq p_0$. In i, form the product $p_x \square r_y \neq 0$. There exists at least one point $t \leqq p_x \square r_y$ and $0 < t_x \leqq p_x$, $0 < t_y \leqq r_y$. Since $t > 0$, $t_x = p_x$, $t_y = p_y$. Hence $[x; y] = [(\bar{t}_x, q); (\bar{t}_y, s)]$, $t_x \neq p_0$, $t_y \neq p_0$. But there exists an element $n \in i$ (Lemma 2b) such that $\pi_n(q) = s$. Thus $[x; y] = [(\bar{t}_x, q); (\bar{t}_y, \pi_n(q))]$ is the map of some point in $(\bar{t})$.

THEOREM 2. *A projective algebra defined on all subsets of a set i of points p is isomorphic to a projective algebra of certain subsets of a direct product $[X; Y]$, with preservation of unit ($i \to [X; Y]$), union, intersection, complement, projection, and $\square$-product.*

The correspondence is that already defined. If now $a \to A$, $b \to B$,

$a \vee b \to S$, $a \wedge b \to P$, one readily verifies that $S = A \cup B$, $P = A \cap B$, since each member includes the other. For the latter equality one uses Lemma 3.

Now let $a \to A$, $a' \to B$. Then $B =$ complement of A in $[X; Y]$. Suppose $[x; y] \in B$, i. e., $s \to [x; y]$ where $s \in a'$. Then $[x; y] \in A'$, for if $[x; y] \in A$ one has $t \to [x; y]$, $t \in a$. By Lemma 3, $s = t$. Hence $B \subset A'$. Now let $[x; y] \in A'$. By Lemma 4, $s \to [x; y]$, for some $s \in i$. Hence $s \in a'$, and $[x; y] \in B$.

Next let $a \to A$, $a_x \to B$. Then $(A)_x \doteq B$. Let $[x; p_0] \in A_x$. For some y, $[x; y] \in A$, hence $p \to [x; y]$ for $p \in a$. By examining the correspondences (a-d) one sees that $p \to [x; y]$ implies $p_x \to [x; p_0]$ in all cases. For example, if p is of the type (d), $p_x \in i_x$, and $p_x \neq p_0$ (for if so, $p \leqq i_y$, cf. (C.18)). Thus $p_x \to [x; p_0]$ under (b). Hence in general $[x; p_0] \in B$. Similarly, if $a \to A$, $a_y \to C$, then $(A)_y = C$.

Finally, let $0 < a \leqq i_x$, $0 < b \leqq i_y$, and $a \to A$, $b \to B$, $a \square b \to C$. Then $A \square B = C$. For, let $[x; y] \in C$. Then $p \to [x; y]$, $p \in a \square b$, and $p_x \leqq a$, $p_y \leqq b$. But as pointed out above, $p_x \to [x; p_0]$ and $p_y \to [p_0; y]$, hence $[x; p_0] \in A$, $[p_0; y] \in B$. Thus $[x; y] \in A \square B$.

Conversely if $[x; y] \in A \square B$, $[x, p_0] \in A$, $[p_0; y] \in B$. Then for $p \leqq a$, $q \leqq b$, $p \to [x; p_0]$, $q \to [p_0; y]$. But then $0 < p \square q \leqq a \square b$, and there is at least one point $t \leqq p \square q \leqq a \square b$. Thus $t_x = p$, $t_y = q$, and $t \to [x; y]$. Hence $[x; y] \in C$.

University of Wisconsin,
and
University of Southern California.

CORRESPONDING TYPE CONTINUED FRACTIONS.*

By Evelyn Frank.[1]

1. Introduction. This paper is concerned with a development of properties of *corresponding type continued fractions*

$$1+\frac{a_1 z^{\alpha_1}}{1}+\frac{a_2 z^{\alpha_2}}{1}+\frac{a_3 z^{\alpha_3}}{1}+\cdots, \tag{1.1}$$

in which the a_p are complex numbers and the α_p are positive integers. For the sake of brevity, we shall call (1.1) a *C-fraction.*

It has been shown by Leighton and Scott [1][2] that an arbitrary power series $P(z)=1+c_1z+c_2z^2+\cdots$ can be expanded into a uniquely determined C-fraction, which terminates if, and only if, the power series represents a rational function of z. Scott and Wall [4] investigated the relationship of the C-fraction to the table of Padé approximants for $P(z)$. They called the C-fraction and its power series *regular* if all its approximants are Padé approximants, and determined a certain class of regular C-fractions.

In the present paper, we continue this investigation. The principal results may be summarized as follows:

A. Algorithm for expanding an arbitrary power series into a C-fraction. We find that there is an algorithm for expanding a power series into a C-fraction which is analogous to that given by Wall [6] for J-fractions. This gives an essential simplification of the expansion problem (cf. [1], p. 598; [4], p. 329), inasmuch as *all long division is eliminated* (**2**).

B. The Padé table. If in the Padé table for a power series $P(z)$ there are two equal approximants, then there must be a square *block* of $(r+1)^2$ equal approximants. Padé ([2], p. 429) determined necessary and sufficient conditions for the *order* r of a block to be zero. We have extended this theorem to cover *arbitrary* orders (Theorem 3.1).

C. Regular C-fractions. From a consideration of the geometrical arrangement of the blocks in the Padé table and their relationship to the exponents

---

* Received May 27, 1945. Presented to the American Mathematical Society, April 28, 1945.

[1] The author wishes to give thanks to Professor H. S. Wall who suggested the topic of this paper and who assisted in its preparation.

[2] Numbers in square brackets refer to the bibliography.

α_p in the C-fraction, we obtain a new formulation of the condition for the C-fraction to be regular (Theorem 4.1). In particular, we obtain the class of *α-regular* C-fractions in which the condition for regularity depends only upon the exponents α_p (Theorem 4.2).

D. Characterization of regular power series in terms of their coefficients. It is well known that a power series is semi-normal, i.e., $\alpha_p = 1$, if, and only if, certain determinants formed from the coefficients of the power series are different from zero. We extend this theorem to cover the class of α-regular power series, and indicate how an analogous theorem may be obtained for regular power series (**5**).

E. Some transformations of C-fractions. A remarkable example of a C-fraction was given by Ramanujan [3], namely,

$$1 + \frac{z}{1} + \frac{z^2}{1} + \frac{z^3}{1} + \frac{z^4}{1} + \cdots.$$

This is identically equal to

$$1 + z - \frac{z^3}{1} + \frac{z^2}{1} - \frac{z}{1} + \frac{z}{1} - \frac{z^3}{1} + \frac{z^3}{1} - \frac{z^2}{1} + \frac{z^2}{1} - \frac{z^4}{1} + \frac{z^4}{1} - \cdots.$$

We find that this transformation of the Ramanujan C-fraction can be carried out for all C-fractions for which $\alpha_2 + \alpha_4 + \cdots + \alpha_{2p+2} > \alpha_3 + \alpha_5 + \cdots + \alpha_{2p+1} > \alpha_2 + \alpha_4 + \cdots + \alpha_{2p}$ (Theorem 6.1). We also obtain a number of other transformations (Theorems 6.2, 6.4, 6.5, and 6.6).

2. Algorithm for expanding a power series into a C-fraction. The following theorem can be used to obtain the C-fraction for an arbitrary power series, and can also be used to expand a C-fraction into a power series.

THEOREM 2.1. *Let $P(z) = 1 + c_1 z + c_2 z^2 + \cdots$ be an arbitrary formal power series, and determine polynomials $B_p(z) = \beta_0^{(p)} + \beta_1^{(p)} z + \beta_2^{(p)} z^2 + \cdots$, numbers $a_p \neq 0$, and positive integers α_p, $p = 1, 2, \cdots$, by means of the recurrence formulas*

$$(2.1) \quad \beta_0^{(0)} = \beta_0^{(1)} = 1, \quad \beta_p^{(0)} = \beta_p^{(1)} = 0 \text{ for } p > 0; \quad \alpha_0 = 0;$$

$$(2.2) \quad (c_n, c_{n-1}, c_{n-2}, \cdots) \begin{pmatrix} \beta_0^{(p)} \\ \beta_1^{(p)} \\ \beta_2^{(p)} \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} = \begin{cases} 0 \text{ if } \alpha_0 + \alpha_1 + \cdots + \alpha_p < n < \alpha_0 + \alpha_1 + \cdots + \alpha_{p+1}, \\ (-1)^p a_1 a_2 \cdots a_{p+1} \text{ if } n = \alpha_0 + \alpha_1 + \cdots + \alpha_{p+1}, \\ (p = 0, 1, 2, \cdots); \end{cases}$$

$$(2.3) \quad B_{p+1}(z) = B_p(z) + a_{p+1} z^{\alpha_{p+1}} B_{p-1}(z), \qquad (p = 1, 2, 3, \cdots).$$

Then

(2.4) $$1+\frac{a_1z^{\alpha_1}}{1}+\frac{a_2z^{\alpha_2}}{1}+\frac{a_3z^{\alpha_3}}{1}+\cdots$$

is the C-fraction for $P(z)$. It terminates if, and only if, $P(z)$ is a rational function of z, in which case there is a value p' of p such that the left-hand member of (2.2) is 0 for all $n > \alpha_0+\alpha_1+\cdots+\alpha_{p'}$. The polynomials $B_p(z)$ are the denominators of (2.4). The numerators are $A_p(z)=\gamma_0^{(p)}+\gamma_1^{(p)}z+\cdots+\gamma_s^{(p)}z^s$ where $s=s_p$ is at most the sum of the $[(p+1)/2]$ largest integers in the set $\alpha_0,\alpha_1,\alpha_2,\cdots,\alpha_p$, and, if $c_0=1$,

(2.5) $$(\gamma_0^{(p)},\gamma_1^{(p)},\gamma_2^{(p)},\cdots)=(\beta_0^{(p)},\beta_1^{(p)},\beta_2^{(p)},\cdots)\begin{pmatrix} c_0, & c_1, & c_2, & \cdots, & c_s, & 0, & \cdots \\ 0, & c_0, & c_1, & \cdots, & c_{s-1}, & 0, & \cdots \\ 0, & 0, & c_0, & \cdots, & c_{s-2}, & 0, & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}.$$

Proof. Suppose first that $P(z)$ is not a rational function of z, so that the recurrent process obviously cannot terminate, and the infinite C-fraction (2.4) can be constructed. From (2.1), (2.3) it follows that $B_p(z)$ is its p-th denominator. Its numerators are given by the recurrence formulas

(2.6) $$A_0(z)=1,\ A_1(z)=1+a_1z^{\alpha_1},\ A_{p+1}(z)=A_p(z)+a_{p+1}z^{\alpha_{p+1}}A_{p-1}(z),\quad (p=1,2,\cdots).$$

One may verify at once that $P(z)B_0(z)-A_0(z)=a_1z^{\alpha_1}+\cdots$. Using induction, we suppose that

(2.7) $$P(z)B_p(z)-A_p(z)=(-1)^pa_1a_2\cdots a_{p+1}z^{\alpha_1+\alpha_2+\cdots+\alpha_{p+1}}+\cdots$$

for $p=0,1,\cdots,m$, and shall prove it for $p=m+1$. We have

$$\begin{aligned} P(z)B_{m+1}(z)-A_{m+1}(z) &= P(z)[B_m(z)+a_{m+1}z^{\alpha_{m+1}}B_{m-1}(z)] \\ &\quad -[A_m(z)+a_{m+1}z^{\alpha_{m+1}}A_{m-1}(z)]=[P(z)B_m(z)-A_m(z)] \\ &\quad +a_{m+1}z^{\alpha_{m+1}}[P(z)B_{m-1}(z)-A_{m-1}(z)]=(z^{1+\alpha_1+\alpha_2+\cdots+\alpha_{m+1}}), \end{aligned}$$

where (z^k) denotes a power series containing no lower power of z than the k-th power. Since, as may be easily seen, the degree of $A_{m+1}(z)$ is at most the sum of the $[(m+2)/2]$ largest integers in $\alpha_0,\alpha_1,\alpha_2,\cdots,\alpha_{m+1}$, and is therefore less than $\alpha_0+\alpha_1+\alpha_2+\cdots+\alpha_{m+1}$, it follows that the terms of degree greater than this value in $P(z)B_{m+1}(z)-A_{m+1}(z)$ are the same as those in $P(z)B_{m+1}(z)$. By (2.2) the first such term which occurs is

$$(-1)^{m+1}a_1a_2\cdots a_{m+2}z^{\alpha_1+\alpha_2+\cdots+\alpha_{m+2}},$$

and consequently (2. 7) holds for $p = m + 1$, as was to be proved. The formula (2. 5) may be obtained by equating coefficients of $z^0, z^1, \cdots, z^s$ on either side of (2. 7).

If $P(z)$ is a rational function of z, the recurrent process must terminate with some $p = p'$, for such a function cannot have an infinite C-fraction. If, in the preceding induction, we assume that $m < p'$, it follows that (2. 7) holds for $p = 0, 1, \cdots, p' - 1$, while $P(z)B_{p'}(z) - A_{p'}(z) \equiv 0$.

To illustrate the theorem, let $P(z) = 1 - z^2 + z^4$, a polynomial. Then we have $a_1 = -1$, $\alpha_1 = 2$, $a_2 = 1$, $\alpha_2 = 2$, so that $B_2(z) = 1 + z^2$. From (2. 2) with $p = 2$, $n > \alpha_1 + \alpha_2 = 4$, we get

$$(0, 1, 0)\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 0, \qquad (0, 0, 1)\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 1 = a_1 a_2 a_3 = -a_3,$$

so that $a_3 = -1$, $\alpha_3 = 2$, $B_3 = 1$, and the process terminates. Hence

$$1 - z^2 + z^4 = 1 - \frac{z^2}{1} + \frac{z^2}{1} - \frac{z^2}{1}.$$

The formulas of Theorem 2. 1 may also be used to obtain the power series expansion of a given C-fraction.

3. A theorem on the Padé table. To the point (m, n) with non-negative integral coordinates m and n, in the cartesian plane, we let correspond the *Padé approximant* $f_{m,n}(z) = N_{m,n}(z)/D_{m,n}(z)$ of the power series $P(z) = c_0 + c_1 z + c_2 z^2 + \cdots$ $(c_0 \neq 0)$, determined uniquely by the following conditions: $D_{m,n} \not\equiv 0$ is of degree $\leq m$, $N_{m,n}(z)$ is of degree $\leq n$, $PD_{m,m} - N_{m,n} = (z^{m+n+1})$. We shall suppose that the positive x-axis extends downward and the positive y-axis extends to the right. We may also regard $f_{m,n}(z)$ as occupying the *square* $[m, n]$ with vertices (m, n), $(m, n+1)$, $(m+1, n)$, $(m+1, n+1)$. This geometrical configuration is called the *Padé table* for $P(z)$ (cf., for instance, [2], chapter 10). A particular rational fraction may occur more than once as a Padé approximant in the table, but in any case it fills a *block* of squares

$$[m + p, n + q], \qquad (p, q = 0, 1, \cdots, r), \tag{3.1}$$

and occurs nowhere else in the table ([2], p. 427). We call r the *order* of the approximant and of the block of squares which it occupies. The theorem which we shall now prove gives necessary and sufficient conditions upon the coefficients of $P(z)$ for the Padé table to contain this block of order r.

THEOREM 3.1. *Let $P(z) = c_0 + c_1 z + c_2 z^2 + \cdots$, $c_0 \neq 0$, be an arbitrary power series, and put*

$$(3.2) \qquad \Delta_{m,n} = \begin{vmatrix} c_{n-m}, & c_{n-m+1}, & \cdots, & c_n \\ c_{n-m+1}, & c_{n-m+2}, & \cdots, & c_{n+1} \\ \cdot & \cdot & \cdot & \cdot \\ c_n, & c_{n+1}, & \cdots, & c_{n+m} \end{vmatrix}, \qquad (m, n = 0, 1, 2, \cdots),$$

where we agree to set $c_p = 0$ if $p < 0$. It will be convenient to define $\Delta_{-1,n} = 1$, $(n = 0, 1, 2, \cdots)$. In order for the Padé table for $P(z)$ to contain the block (3.1) of order r, it is necessary and sufficient that all of the following conditions hold:

$$(3.3) \quad \begin{array}{lll} \text{(i) } \Delta_{t-1,s} \neq 0, & \text{(ii) } \Delta_{t-1,s+1} \neq 0, & \text{(iii) } \Delta_{t,s} \neq 0, \\ \text{(iv) } \Delta_{t+k,s+k+1} = 0, & (k = 0, 1, \cdots, r-1), & \text{(v) } \Delta_{t+r,s+r+1} \neq 0. \end{array}$$

Proof. It is clearly necessary and sufficient for the Padé table to contain the given block that

$$(3.4) \quad \begin{array}{lll} \text{(i}') \ f_{t-1,s-1} \neq f_{t,s}, & \text{(ii}') \ f_{t-1,s} \neq f_{t,s}, & \text{(iii}') \ f_{t,s-1} \neq f_{t,s}, \\ \text{(iv}') \ f_{t+p,s+p} = f_{t,s}, & (p = 1, 2, \cdots, r), & \text{(v}') \ f_{t+r+1,s+r+1} \neq f_{t,s}. \end{array}$$

We shall show first that (3.3) is *necessary* for (3.4). Let $N = \alpha_0 + \alpha_1 z + \cdots + \alpha_s z^s$, $D = \beta_0 + \beta_1 z + \cdots + \beta_t z^t$,

$$\Gamma_{m,n} = \begin{pmatrix} c_n, & c_{n-1}, & \cdots, & c_{n-m} \\ c_{n+1}, & c_n, & \cdots, & c_{n-m+1} \\ \cdot & \cdot & \cdot & \cdot \\ c_{n+m-1}, & c_{n+m-2}, & \cdots, & c_{n-1} \end{pmatrix}, \qquad \alpha = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}.$$

Then the condition for N/D to be the Padé approximant $f_{t,s}(z)$ of $P(z)$ is:

$$(3.5) \qquad \Gamma_{s+1,0}\beta = \alpha, \qquad \Gamma_{t,s+1}\beta = 0, \quad \beta \neq 0.$$

If $\Delta_{t-1,s} = 0$, then (3.5) has a solution with $\beta_0 = 0$, $\alpha_0 = 0$, so that $f_{t-1,s-1} = f_{t,s}$; if $\Delta_{t-1,s+1} = 0$, there is a solution with $\beta_t = 0$, so that $f_{t-1,s} = f_{t,s}$; if $\Delta_{t,s} = 0$, there is a solution with $\alpha_s = 0$, so that $f_{t,s-1} = f_{t,s}$. Thus (i), (ii), (iii) are necessary for (i′), (ii′), (iii′), respectively. In order for (iv′) to hold, it is necessary that z^pN/z^pD be a Padé approximant for $p = 1, 2, \cdots, r$, i.e., that $Pz^pD - z^pN$ contain no lower power of z than the $(t + s + 2p - 1)$-th, for $p = 1, 2, \cdots, r$. If $\beta^{(p)}$ is the matrix obtained

from β by inserting p zeros above β_0, this condition can be written $\Gamma_{t+p,s+p+1}\beta^{(p)} = 0$, $(p = 1, 2, \cdots, r)$. Inasmuch as $\beta^{(p)} \neq 0$, this implies (iv). Finally, (v) is necessary for (v′), for the same reason that (i) is necessary for (i′).

We shall now prove that (3. 3) is *sufficient* for (3. 4). By (i) the system (3. 5) has a solution α, β, and every other solution has the form $c\alpha$, $c\beta$, where c is a constant. Moreover, $\beta_0 \neq 0$. Therefore, the fraction N/D is *irreducible*. For, if $N/D \equiv N^*/D^*$ where N^* and D^* are of lower degree by $h > 0$ than N and D, then $Pz^hD^* - z^hN^*$ would contain no lower power of z than the $(s + t + 1)$-th ([2], p. 423). Hence, the polynomials z^hD^* and z^hN^* could be determined by solving (3. 5). But in this solution we would have $\beta_0 = 0$, which is impossible. Condition (ii) implies that $\beta_t \neq 0$. This together with the fact that N/D is irreducible shows that (ii′) holds. Condition (iii) implies that $\alpha_s \neq 0$, so that (iii′) holds. To prove (iv′), let $f_{t+p,s+p} = N^{(p)}/D^{(p)}$, and let $\alpha^{(p)}$, $\beta^{(p)}$ be the one-column matrices of the coefficients of $N^{(p)}$ and $D^{(p)}$. Since $PD^{(p)} - N^{(p)} = (z^{s+t+2p-1})$, we have

$$(3.\ 6) \qquad \Gamma_{s+p+1,0}\beta^{(p)} = \alpha^{(p)}, \quad \Gamma_{t+p;s+p+1}\beta^{(p)} = 0, \qquad (p = 1, 2, \cdots, r).$$

Since $\Delta_{t,s+1} = 0$, we see that for $p = 1$ (3. 6) has the solution $\beta_0^{(1)} = 0$, $\beta_k^{(1)} = \beta_{k-1}$, $(k = 1, 2, \cdots, t + 1)$, $\alpha_0^{(1)} = 0$, $\alpha_k^{(1)} = \alpha_{k-1}$, $(k = 1, 2, \cdots, s + 1)$. Hence it follows that $N^{(1)}/D^{(1)} \equiv N/D$, and (iv′) holds for $p = 1$. Then, since $\Delta_{t+1,s+2} = 0$, it follows that, for $p = 2$, (3. 6) has the solution $\beta_0^{(2)} = 0$, $\beta_k^{(2)} = \beta_{k-1}^{(1)}$, $(k = 1, 2, \cdots, t + 2)$, $\alpha_0^{(2)} = 0$, $\alpha_k^{(2)} = \alpha_{k-1}^{(1)}$, $(k = 1, 2, \cdots, s + 2)$, where $\alpha^{(1)}$, $\beta^{(1)} \neq 0$ is a solution of (3. 6) for $p = 1$. Hence $N^{(2)}/D^{(2)} \equiv N^{(1)}/D^{(1)} \equiv N/D$, so that (iv′) holds for $p = 2$. Continuing this process we conclude that (iv′) holds for $p = 1, 2, \cdots, r$. Finally, since $\Delta_{t+r,s+r+1} \neq 0$, $PD^{(r)} - N^{(r)}$ actually contains the $(t + s + 2r + 1)$-th power of z, so that (v′) holds. For a like reason, (i) implies (i′). This completes the proof of Theorem 3. 1.

The preceding proof contains the proof of the following theorem.

THEOREM 3. 2. *In the Padé table for $P(z)$, we have for some integers p, q, r:*

$$(3.\ 7) \quad \text{(i) } f_{q-1,p-1} \neq f_{q,p}, \ \text{(ii) } f_{q+k,p+k} = f_{q,p}, \ (k = 1, 2, \cdots, r), \ \text{(iii) } f_{q+r+1,p+r+1} \neq f_{q,p}$$

if, and only if,

$$(3.\ 8) \quad \text{(i}'\text{) } \Delta_{q-1,p} \neq 0, \ \text{(ii}'\text{) } \Delta_{q+k,p+k+1} = 0, \ (k = 0, 1, \cdots, r - 1), \ \text{(iii}'\text{) } \Delta_{q+r,p+r+1} \neq 0.$$

4. Regular C-fractions. The C-fraction (1. 1) and its power series $P(z) = 1 + c_1z + c_2z^2 + \cdots$ are called *regular* if every approximant

$A_p(z)/B_p(z)$ is a Padé approximant for $P(z)$. We shall first reformulate some results of Scott and Wall [4] on the Padé table for a regular power series. The novel aspect of this formulation consists in the introduction of the *α-polygon*. We shall then characterize regular C-fractions in terms of geometrical properties of the Padé table, and then (5) in terms of the coefficients of the power series.

If s_p and t_p are the degrees of $A_p(z)$ and $B_p(z)$, respectively, then $P(z)$ is regular if, and only if, for every p, $P(z)B_p(z) - A_p(z) = (z^{s_p+t_p+1})$ or if, and only if, there is an integer r_p such that

$$(4.1)\quad s_p + t_p + r_p + 1 = \alpha_1 + \alpha_2 + \cdots + \alpha_{p+1},\ r_p \geq 0,\ (p = 0, 1, 2, \cdots).$$

Since the approximant is irreducible, r_p is its order [2]. Let Q_p denote the block of order r_p whose squares are occupied by $A_p(z)/B_p(z)$. The coordinates of its vertices are (t_p, s_p), $(t_p, s_p + r_p + 1)$, $(t_p + r_p + 1, s_p)$, $(t_p + r_p + 1, s_p + r_p + 1)$. For all points on the diagonal connecting the second and third of these vertices, the sum of the coordinates is $s_p + t_p + r_p + 1$. By (4.1), this sum increases with increasing p. Hence, as Q_p and Q_{p+1} do not overlap, it follows that for any given p either

$$(4.2)\qquad s_{p+1} > s_p + r_p,$$
or
$$(4.3)\qquad t_{p+1} > t_p + r_p.$$

By (2.3), (2.6), it follows that

$$(4.4)\qquad s_{p+1} = s_{p-1} + \alpha_{p+1},$$
or
$$(4.5)\qquad t_{p+1} = t_{p-1} + \alpha_{p+1},$$

according as (4.2) or (4.3) holds, respectively. Hence,

$$(4.6)\qquad s_{p+1} - (s_p + r_p) = t_p - (t_{p-1} + r_{p-1}) > 0,$$
or
$$(4.7)\qquad t_{p+1} - (t_p + r_p) = s_p - (s_{p-1} + r_{p-1}) > 0,$$

respectively. One may then readily show by mathematical induction that

$$(4.8)\qquad \begin{array}{ll} s_{2p+1} > s_{2p} + r_{2p}, & s_{2p+2} \leq s_{2p+1} + r_{2p+1}, \\ t_{2p+1} \leq t_{2p} + r_{2p}, & t_{2p+2} > t_{2p+1} + r_{2p+1}, \end{array}\qquad (p = 0, 1, 2, \cdots).$$

Hence we find, by (4.4), (4.5) that

$$(4.9)\qquad \begin{array}{l} s_{2p-1} = \alpha_1 + \alpha_3 + \cdots + \alpha_{2p-1}, \\ t_{2p} \ = \alpha_2 + \alpha_4 + \cdots + \alpha_{2p}, \end{array}\qquad (p = 1, 2, 3, \cdots).$$

By (4.1) it follows that

$$(4.10)\qquad \begin{aligned} 1 + s_{2p} + r_{2p} &= \alpha_1 + \alpha_3 + \cdots + \alpha_{2p+1}, \\ 1 + t_{2p-1} + r_{2p-1} &= \alpha_2 + \alpha_4 + \cdots + \alpha_{2p}, \end{aligned}$$

or

$$(4.11)\qquad \begin{aligned} s_{2p+1} &= s_{2p} + r_{2p} + 1, \\ t_{2p} &= t_{2p-1} + r_{2p-1} + 1. \end{aligned}$$

These results have a simple geometrical interpretation in the Padé table. We first draw the α-polygon $OV_0V_1V_2\cdots$ with vertices $O = (0, 0)$, $V_0 = (0, \alpha_1)$, $V_{2p-1} = (\alpha_2 + \alpha_4 + \cdots + \alpha_{2p}, \alpha_1 + \alpha_3 + \cdots + \alpha_{2p-1})$, $V_{2p} = (\alpha_2 + \alpha_4 + \cdots + \alpha_{2p}, \alpha_1 + \alpha_3 + \cdots + \alpha_{2p+1})$, $(p = 1, 2, 3, \cdots)$. By (4.9), (4.10), the vertices of Q_{2p} are

$$W_{2p} = (\alpha_2 + \alpha_4 + \cdots + \alpha_{2p}, s_{2p}), \quad V_{2p},$$
$$T_{2p} = (t_{2p} + r_{2p} + 1, s_{2p}), \qquad U_{2p} = (t_{2p} + r_{2p} + 1, \alpha_1 + \alpha_3 + \cdots + \alpha_{2p+1}),$$

and the vertices of Q_{2p+1} are

$$U_{2p+1} = (t_{2p+1}, \alpha_1 + \alpha_3 + \cdots + \alpha_{2p+1}), \quad T_{2p+1} = (t_{2p+1}, s_{2p+1} + r_{2p+1} + 1),$$
$$V_{2p+1}, \qquad W_{2p+1} = (\alpha_2 + \alpha_4 + \cdots + \alpha_{2p}, s_{2p+1} + r_{2p+1} + 1).$$

Thus, the α-polygon forms part of the boundary of every square Q_p.

By (4.8) we see that Q_p and Q_{p+1} have a line segment D_p of the α-polygon as common boundary (Fig. 1).

Let $L(\omega)$ denote the straight line whose equation is $y = x + \omega$. Then it is clear that there exist one or more integers ω_p such that the lines $L(\omega_p)$ and $L(1 + \omega_p)$ cut the line segment D_{2p}. The principal diagonal of Q_{2p} lies on or below $L(\omega_p)$, and the principal diagonal of Q_{2p+1} lies on or above $L(1 + \omega_p)$. Hence we see that W_{2p} must lie on or below $L(\omega_p)$, U_{2p+1} must lie on or above $L(1 + \omega_p)$, V_{2p} must lie on or above $L(1 + \omega_p)$, and V_{2p+1} must lie on or below $L(\omega_p)$. These statements are equivalent to the following inequalities:

$$(4.12)\qquad \begin{aligned} &\text{(i)} && s_{2p} \leq t_{2p} + \omega_p, \\ &\text{(ii)} && s_{2p+1} \geq t_{2p+1} + \omega_p + 1, \\ &\text{(iii)} && s_{2p+1} \geq t_{2p} + \omega_p + 1, \\ &\text{(iv)} && s_{2p+1} \leq t_{2p+2} + \omega_p, \qquad (p = 0, 1, 2, \cdots). \end{aligned}$$

We shall now prove that the *necessary* conditions (4.12) and (4.9) are *sufficient* for regularity. We suppose that (4.9) holds, and that integers $\omega_0, \omega_1, \cdots$ exist such that (4.12) holds. Then $s_{2p} + t_{2p} = s_{2p} + \alpha_2 + \alpha_4$

$+\cdots+\alpha_{2p} \leq t_{2p}+\omega_p+\alpha_2+\alpha_4+\cdots+\alpha_{2p} \leq s_{2p+1}-1+\alpha_2+\alpha_4+\cdots+\alpha_{2p}$, or $s_{2p}+t_{2p}+1 \leq \alpha_1+\alpha_2+\cdots+\cdots+\alpha_{2p+1}$. Similarly, $s_{2p+1}+t_{2p+1}=\alpha_1+\alpha_3+\cdots+\alpha_{2p+1}+t_{2p+1} \leq \alpha_1+\alpha_3+\cdots+\alpha_{2p+1}+s_{2p+1}$

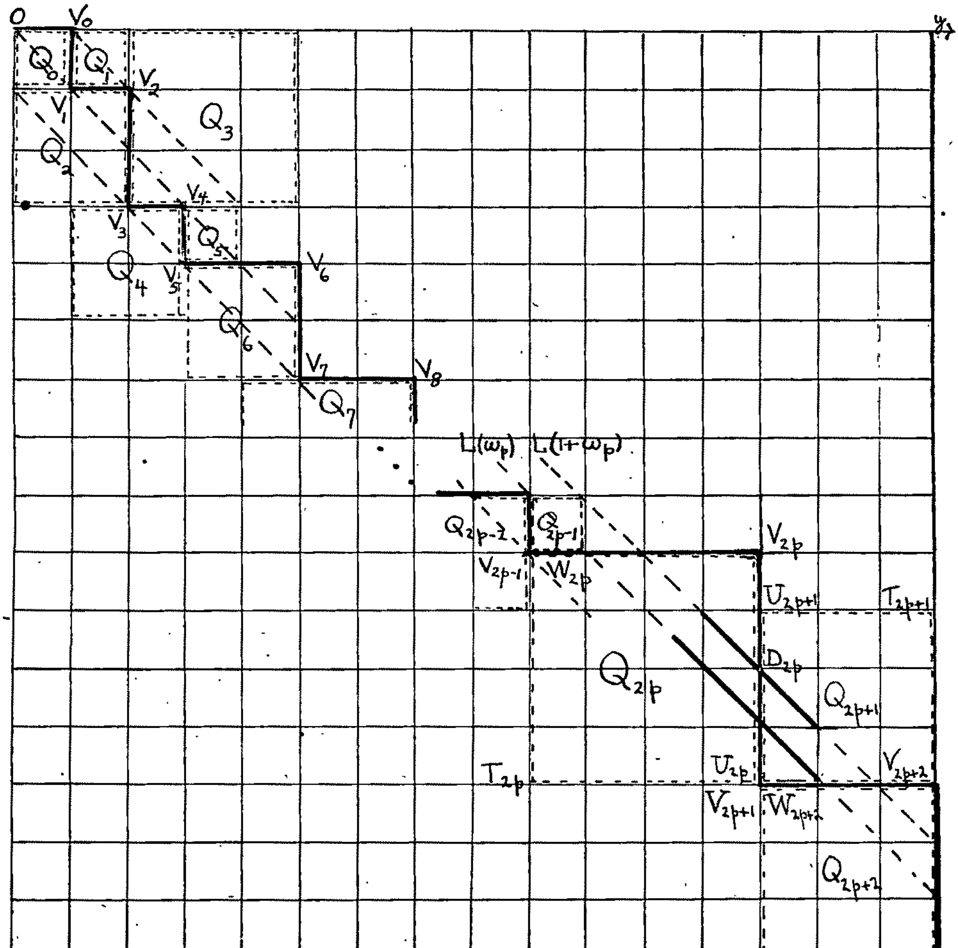


Fig. 1

$-\omega_p-1 \leq \alpha_1+\alpha_3+\cdots+\alpha_{2p+1}+t_{2p+2}-1$, or $s_{2p+1}+t_{2p+1}+1 \leq \alpha_1+\alpha_2+\cdots+\alpha_{2p+2}$. Consequently, (4.1) holds, and hence the C-fraction is regular.

We have proved the following theorem:

THEOREM 4.1. *The C-fraction* (1.1) *is regular if, and only if,* (4.9) *and* (4.12) *hold, where* $\omega_0, \omega_1, \cdots$ *are integers.*

If (4.9) holds, and there is an integer $\omega \geq 0$ such that the straight lines

$L(\omega)$ and $L(1+\omega)$ intersect all the line segments D_p, $(p=0,1,2,\cdots)$, then the C-fraction and its power series are called *α-regular*. We shall prove the following theorem:

THEOREM 4.2. *The C-fraction (1.1) is α-regular if and only if there exists an integer $\omega \geq 0$ such that*

$$(4.13)\quad \begin{aligned} &\alpha_1+\alpha_3+\cdots+\alpha_{2p+1} \geq \alpha_0+\alpha_2+\alpha_4+\cdots+\alpha_{2p}+\omega+1, \\ &\alpha_2+\alpha_4+\cdots+\alpha_{2p+2} \geq \alpha_1+\alpha_3+\cdots+\alpha_{2p+1}-\omega, \quad (p=0,1,2,\cdots), \end{aligned}$$

where $\alpha_0=0$.

Proof. If (1.1) is α-regular, then (4.12) holds with $\omega=\omega_p$, $(p=0,1,2,\cdots)$. Hence, by (4.9) and (4.12) (iii), (iv), we conclude that (4.13) holds.

If, conversely, (4.13) holds, then we shall prove by induction that

$$(4.14)\quad \begin{aligned} &s_{2p}-\omega \leq t_{2p}=\alpha_0+\alpha_2+\alpha_4+\cdots+\alpha_{2p}, \qquad (p=0,1,2,\cdots), \\ &\omega+1+t_{2p-1} \leq s_{2p-1}=\alpha_1+\alpha_3+\cdots+\alpha_{2p-1}, \quad (p=1,2,3,\cdots), \end{aligned}$$

from which it will follow immediately that (1.1) is α-regular. The first relation (4.14) holds for $p=0$ since $\omega \geq 0$. Since $s_1=\alpha_1$, $t_1=0$, $s_2 \leq \max(\alpha_1,\alpha_2)$, $t_2=\alpha_2$, we readily verify (4.14) for $p=1$. Assuming that the relations (4.14) hold for $p \leq n$, $n \geq 1$, we shall prove them for $p=n+1$. From (2.6) we see that $s_{2n+1}=\max(s_{2n}, s_{2n-1}+\alpha_{2n+1})$ if $s_{2n} \neq s_{2n-1}+\alpha_{2n+1}$. But, by our assumption, $s_{2n} \leq \omega+\alpha_2+\alpha_4+\cdots+\alpha_{2n}$ and, by (4.13), $s_{2n-1}+\alpha_{2n+1}=\alpha_1+\alpha_3+\cdots+\alpha_{2n+1}>\alpha_2+\alpha_4+\cdots+\alpha_{2n}+\omega$, so that $s_{2n+1}=\alpha_{2n+1}+s_{2n-1}=\alpha_1+\alpha_3+\cdots+\alpha_{2n+1}$. By (2.3), $t_{2n+1} \leq \max(t_{2n}, t_{2n-1}+\alpha_{2n+1})$, so that

$$\omega+1+t_{2n+1} \leq \left\{ \begin{matrix} \omega+1+t_{2n} \\ \omega+1+t_{2n-1}+\alpha_{2n+1} \end{matrix} \right\} \leq \alpha_1+\alpha_3+\cdots+\alpha_{2n+1}=s_{2n+1}.$$

Next, since $t_{2n+2}=\max(t_{2n+1}, t_{2n}+\alpha_{2n+2})$ provided $t_{2n+1} \neq t_{2n}+\alpha_{2n+2}$, we conclude that $t_{2n+2}=\alpha_2+\alpha_4+\cdots+\alpha_{2n+2}$, inasmuch as, by what we have just proved and (4.13), $t_{2n+1} \leq \alpha_1+\alpha_3+\cdots+\alpha_{2n+1}-\omega-1<\alpha_2+\alpha_4+\cdots+\alpha_{2n+2}=t_{2n}+\alpha_{2n+2}$. Finally, we have

$$s_{2n+2} \leq \left\{ \begin{matrix} s_{2n+1} \\ s_{2n}+\alpha_{2n+2} \end{matrix} \right\} \leq \alpha_2+\alpha_4+\cdots+\alpha_{2n+2}+\omega=t_{2n+2}+\omega.$$

The proof of Theorem 4.2 is now complete.

In general, the regularity of (1.1) depends upon the coefficients a_p as well as upon the exponents α_p. For example, the C-fraction

$$1+\frac{z}{1}-\frac{z}{1}+\frac{z}{1}-\frac{z^2}{1}-\frac{z}{1}+\frac{z}{1}+\frac{z^2}{1}-\frac{z^2}{1}+\frac{z^2}{1}+\frac{z^2}{1}+\frac{z^2}{1}+\cdots$$

is regular (Fig. 1), but if the minus signs are replaced by plus signs, the resulting C-fraction is not regular. The conditions (4.13) involve *only* the exponents α_p, so that α-regularity does not depend upon the coefficients a_p. If, in particular, $\alpha_1 = 1$, and (4.13) holds with $\omega = 0$, then the C-fraction is "absolutely regular" (Scott and Wall [4]). If $\alpha_p = 1$, $(p = 1, 2, 3, \cdots)$, then (4.13) holds with $\omega = 0$. In this case the C-fraction and its power series are called *semi-normal.*

Let us return for a moment to the blocks Q_p in the Padé table for a regular power series $P(z)$. The order r_p of Q_p expresses in a certain sense the amount of the degree of approximation of $A_p(z)/B_p(z)$ to $P(z)$ *in excess of normal.* It is therefore of interest to have upper and lower bounds for the numbers r_p. From (4.9) and (4.12) we readily obtain the inequalities

$$\begin{aligned}
&\alpha_1+\alpha_3+\cdots+\alpha_{2p+1}-1\\
&\qquad\geq r_{2p}\geq(\alpha_1+\alpha_3+\cdots+\alpha_{2p+1}-1)-(\alpha_2+\alpha_4+\cdots+\alpha_{2p}+\omega_p),\\
&\alpha_2+\alpha_4+\cdots+\alpha_{2p+2}-1\\
&\qquad\geq r_{2p+1}\geq(\alpha_2+\alpha_4+\cdots+\alpha_{2p+2}+\omega_p)-(\alpha_1+\alpha_3+\cdots+\alpha_{2p+1}).
\end{aligned}$$

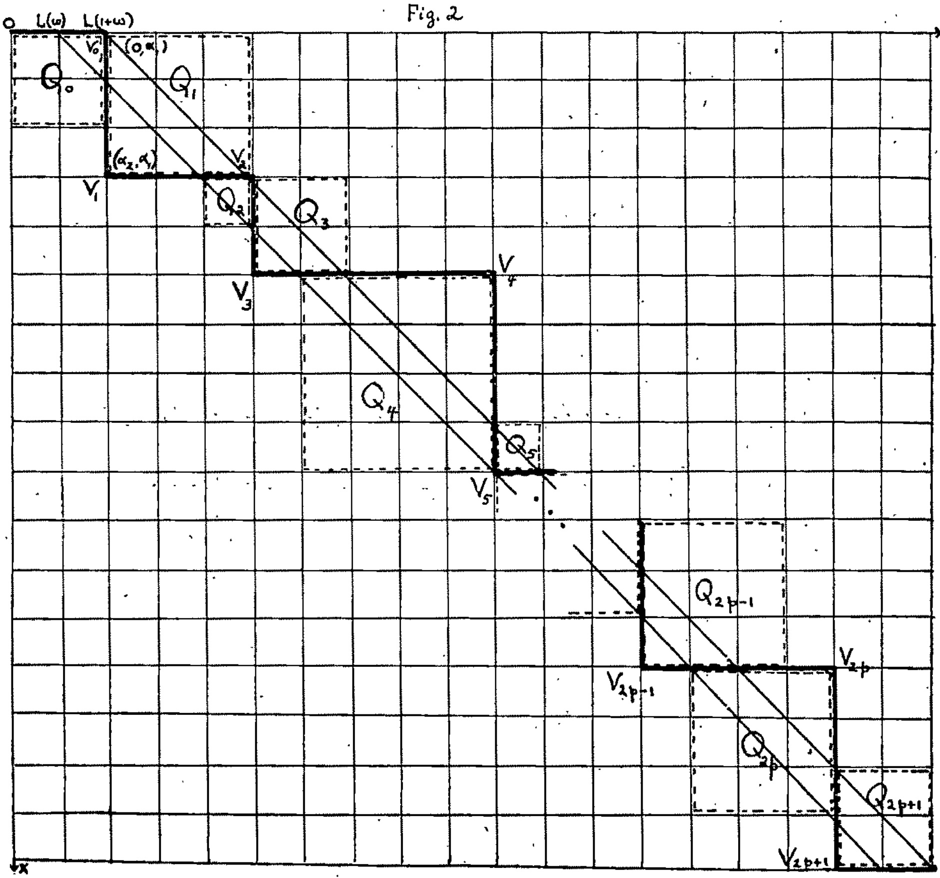
In certain cases, one can determine the exact values of the r_p and hence the exact sizes of the blocks Q_p. This depends upon whether equality holds in (4.12) (i), (ii). For example, if (a) $\alpha_1 = 1$, $\omega_p = 0$, and the a_p are real and positive or actual inequality holds in (4.12) (iii), (iv), or (b) $\omega_p = \alpha_1 - 1$, $\alpha_1 \neq 1$, equality holds in (4.12) (iv) and the a_p are real and positive, then equality holds in (4.12) (i), (ii).

5. Characterization of regular power series in terms of their coefficients. It is well known that $P(z) = 1 + c_1 z + c_2 z^2 + \cdots$ is semi-normal if and only if certain determinants formed from the coefficients c_p are different from zero ([2], p. 304). If $P(z)$ is absolutely regular, Scott and Wall [4] showed that certain of these same determinants are different from zero, but they did not obtain the complete characterization. We shall now proceed to do this for the more general α-regular power series.

Let $P(z)$ be α-regular. Then, in the Padé table, the blocks $Q_0, Q_1, \cdots$ are all cut by the straight lines $L(\omega)$ and $L(1+\omega)$, for a suitable integer $\omega \geq 0$ (see Fig. 2). The line $L(\omega)$ passes through the file of approximants $f_{0,\omega}, f_{1,1+\omega}, f_{2,2+\omega}, \cdots$, and $L(1+\omega)$ passes through $f_{0,1+\omega}, f_{1,2+\omega}, f_{2,3+\omega}, \cdots$. All the even approximants of the C-fraction are contained in the first of these files while all the odd approximants are in the second. The approximants on

$L(\omega)$ which occupy squares in Q_{2p} are equal to one another and are different from those occupying squares outside Q_{2p}. Similarly, the approximants on $L(1+\omega)$ occupying squares in Q_{2p+1} are equal to one another, and are different from those outside Q_{2p+1}. Hence if we put

Fig. 2



$$(5.1)\quad g_0 = 0,\quad g_p = \alpha_2 + \alpha_4 + \cdots + \alpha_{2p},\quad h_p = \alpha_1 + \alpha_3 + \cdots + \alpha_{2p-1},\quad (p = 1, 2, 3, \cdots),$$

we have

$$f_{0,\omega} = f_{1,1+\omega} = \cdots = f_{h_1-\omega-1,h_1-1} \neq f_{h_1-\omega,h_1};$$
$$f_{h_p-\omega-2,h_p-1} \neq f_{h_p-\omega-1,h_p} = f_{h_p-\omega,h_p+1} = \cdots = f_{g_p-1,g_p+\omega} \neq f_{g_p,g_p+\omega+1};$$
$$f_{g_p-1,g_p+\omega-1} \neq f_{g_p,g_p+\omega} = f_{g_p+1,g_p+\omega+1} = \cdots = f_{h_{p+1}-\omega-1,h_{p+1}-1} \neq f_{h_{p-1}-\omega,h_{p+1}},$$
$$(p = 1, 2, 3, \cdots).$$

We may now apply Theorem 3.2 to obtain the following *necessary* conditions for $P(z)$ to be α-regular:

$$\begin{aligned}
&\Delta_{i,\omega+1+i} = 0, \quad (i = 0, 1, \cdots, h_1 - \omega - 2), \quad \Delta_{h_1-\omega-1,h_1} \neq 0;\\
(5.2)\quad &\Delta_{h_p-\omega-2,h_p} \neq 0, \quad \Delta_{h_p-\omega-1+i,h_p+1+i} = 0, \quad (i = 0, 1, \cdots, g_p + \omega - h_p - 1),\\
&\Delta_{g_p-1,g_p+\omega+1} \neq 0\\
&\Delta_{g_p-1,g_p+\omega} \neq 0, \quad \Delta_{g_p+i,g_p+\omega+1+i} = 0, \quad (i = 0, 1, \cdots, h_{p+1} - g_p - 2 - \omega),\\
&\Delta_{h_{p+1}-1-\omega,\, h_{p+1}} \neq 0, \quad (p = 1, 2, 3, \cdots).
\end{aligned}$$

If we change the notation by writing $\Phi(n, \omega) = \Delta_{n-1,n+\omega}$, $\Psi(n, \omega) = \Delta_{n-2,n+\omega}$, we obtain from (5.2) the conditions

$$\begin{aligned}
&\Phi(g_p, \omega) \neq 0, \qquad \Phi(g_p + i, \omega) = 0, \ (i = 1, 2, \cdots, h_{p+1} - g_p - 1 - \omega),\\
(5.3)\quad &\Phi(h_{p+1} - \omega, \omega) \neq 0, \ (p = 0, 1, 2, \cdots);\\
&\Psi(h_p - \omega, \omega) \neq 0, \ \Psi(h_p + i - \omega, \omega) = 0, \ (i = 1, 2, \cdots, g_p + \omega - h_p),\\
&\Psi(g_p + 1, \omega) \neq 0, \ (p = 1, 2, 3, \cdots).
\end{aligned}$$

We find by a process similar to that used by Scott and Wall [4] that the coefficients a_p in the C-fraction are given by the formulas

$$\begin{aligned}
&a_1 = \frac{(-1)^{\alpha_1-1-\omega}\Phi(h_1 - \omega, \omega)}{\Psi(h_1 - \omega, \omega)}, \qquad a_2 = \frac{(-1)^{\alpha_2}\Psi(1 + g_1, \omega)}{\Phi(g_1, \omega)},\\
(5.4)\quad &a_{2p-1} = \frac{(-1)^{\alpha_{2p-1}}\Phi(h_p - \omega, \omega)\Psi(h_{p-1} - \omega, \omega)}{\Psi(h_p - \omega, \omega)\Phi(h_{p-1} - \omega, \omega)},\\
&a_{2p} = \frac{(-1)^{\alpha_{2p}}\Psi(1 + g_p, \omega)\Phi(g_{p-1}, \omega)}{\Psi(1 + g_{p-1}, \omega)\Phi(g_p, \omega)}, \qquad (p = 2, 3, 4, \cdots).
\end{aligned}$$

These values of the a_p may also be obtained from the algorithm (2.2), (2.5).

We note that, by (4.13), the numbers g_p, h_p satisfy the inequalities

$$\begin{aligned}
(5.5)\qquad &\omega + 1 \leq h_1 \leq g_1 + \omega \leq h_2 - 1\\
&\qquad \leq g_2 + \omega - 1 \leq h_3 - 2 \leq g_3 + \omega - 2 \leq \cdots.
\end{aligned}$$

Moreover, if g_p and h_p are any numbers satisfying (5.5), then the C-fraction with exponents

$$(5.6)\qquad \alpha_{2p} = g_p - g_{p-1}, \quad \alpha_{2p-1} = h_p - h_{p-1}, \quad (g_0 = h_0 = 0),$$

is α-regular.

We shall now prove that the conditions (5.3), which are necessary for α-regularity, are also *sufficient*. We suppose that (5.3) holds, where the g_p and h_p are any numbers satisfying (5.5), for some non-negative integer ω. We then determine numbers a_p by the formulas (5.4) and numbers α_p by (5.6), and form the C-fraction (1.1), which, by Theorem 4.2, is necessarily α-regular. This C-fraction has a power series expansion $P'(z) = 1 + c'_1 z + c'_2 z^2 + \cdots$. It is required to show that $c'_p = c_p$, $(p = 1, 2, 3, \cdots)$. From the Padé table for $P'(z)$ we see by what has already been proved that

(5.3) and (5.4) hold with the c_p replaced by the c'_p. It is then clear that the a_p may be determined from the algorithm of 2 if we start either with $P(z)$ or with $P'(z)$. Inasmuch as that algorithm determines the c_p uniquely in terms of the a_p for a given set of the α_p, it follows that $c'_p = c_p$, $(p = 1, 2, 3, \cdots)$.

We have proved the following theorem.

THEOREM 5.1. *A power series $P(z) = 1 + c_1 z + c_2 z^2 + \cdots$ is α-regular if and only if there exist positive integers $\alpha_1, \alpha_2, \alpha_3, \cdots$ and an integer $\omega \geq 0$ such that* (4.13) *holds and such that the coefficients c_p satisfy the conditions* (5.3), *where the g_p and h_p are defined by* (5.1).

As a corollary we have the well-known theorem that $P(z)$ is semi-normal if and only if the determinants $\Phi(p, 0)$, $\Psi(p, 0)$, $(p = 1, 2, 3, \cdots)$, are all different from zero.

Conditions for *regularity* similar to (5.2) can be set up in terms of the coefficients c_p, the parameters ω_p, and additional parameters ω'_p. The latter determine diagonal lines $L(\omega'_p)$ and $L(1 + \omega'_p)$ which intersect the common boundaries of Q_{2p+1} and Q_{2p+2}. Formulas analogous to (5.4) for the a_p may also be obtained. As these conditions and formulas are somewhat complicated, we shall omit them here.

6. Some transformations of *C*-fractions. Let $P(z) = 1 + c_1 z + c_2 z^2 + \cdots$ be a power series which does not represent a rational function of z, and let (1.1) be its C-fraction expansion. Let c_k be the first of the coefficients $c_1, c_2, c_3, \cdots$ which is not zero. Then the power series

$$(6.1) \quad P^*(z) = \frac{P(z) - 1}{c_k z^k}, \quad Q(z) = \frac{1}{P(z)}, \quad P_1(z) = P(z) + Cz^{\alpha_1},$$

where C is a constant, do not represent rational functions of z, and must therefore have non-terminating C-fraction expansions. There are a number of cases where the coefficients and exponents in the C-fractions for the power series (6.1) can be expressed in a simple way in terms of the coefficients and exponents in the C-fraction for $P(z)$. The problem is clearly equivalent to the problem of finding the coefficients b_p, d_p, e_p and exponents β_p, δ_p, ϵ_p in the three continued fractions

$$(6.2) \qquad 1 + a_1 z^{\alpha_1} + \frac{b_1 z^{\beta_1}}{1} + \frac{b_2 z^{\beta_2}}{1} + \frac{b_3 z^{\beta_3}}{1} + \cdots,$$

$$(6.3) \qquad \frac{1}{1} + \frac{d_1 z^{\delta_1}}{1} + \frac{d_2 z^{\delta_2}}{1} + \frac{d_3 z^{\delta_3}}{1} + \cdots,$$

and,

$$(6.4) \qquad 1 - Cz^{\alpha_1} + \frac{e_1 z^{\epsilon_1}}{1} + \frac{e_2 z^{\epsilon_2}}{1} + \frac{e_3 z^{\epsilon_3}}{1} + \cdots,$$

which are all equal to (1.1) in the sense that the power series for (6.2), (6.3), (6.4), and (1.1) are all identical with one another. We shall obtain the required formulas in a number of cases.

THEOREM 6.1. *If the exponents α_p of (1.1) satisfy the inequalities*

$$(6.5) \quad \alpha_2 + \alpha_4 + \cdots + \alpha_{2p+2} > \alpha_3 + \alpha_5 + \cdots + \alpha_{2p+1} > \alpha_2 + \alpha_4 + \cdots + \alpha_{2p}, \quad (p = 1, 2, 3, \cdots),$$

then the coefficients b_p and exponents β_p in (6.2) are given by the formulas

$$b_1 = -a_1 a_2,\ b_2 = a_2,\ \beta_1 = \alpha_1 + \alpha_2,\ \beta_2 = \alpha_2,$$

$$(6.6) \quad b_{4p-1} = -b_{4p} = -\frac{a_3 a_5 \cdots a_{2p+1}}{a_2 a_4 \cdots a_{2p}}, \quad b_{4p+1} = -b_{4p+2} = -\frac{a_2 a_4 \cdots a_{2p+2}}{a_3 a_5 \cdots a_{2p+1}},$$

$$\beta_{4p-1} = \beta_{4p} = (\alpha_3 + \alpha_5 + \cdots + \alpha_{2p+1}) - (\alpha_2 + \alpha_4 + \cdots + \alpha_{2p}),$$

$$\beta_{4p+1} = \beta_{4p+2} = (\alpha_2 + \alpha_4 + \cdots + \alpha_{2p+2}) - (\alpha_3 + \alpha_5 + \cdots + \alpha_{2p+1}), \quad (p = 1, 2, 3, \cdots).$$

Proof. We take the even part of (6.2), i.e., we form the C-fraction whose approximants are the even approximants of (6.2) ([2], p. 201):

$$(6.7) \quad 1 + a_1 z^{\alpha_1} + \frac{b_1 z^{\beta_1}}{1 + b_2 z^{\beta_2}} - \frac{b_2 b_3 z^{\beta_2 + \beta_3}}{1 + b_3 z^{\beta_3} + b_4 z^{\beta_4}} - \frac{b_4 b_5 z^{\beta_4 + \beta_5}}{1 + b_5 z^{\beta_5} + b_6 z^{\beta_6}} - \cdots.$$

We now determine the b_p and β_p by the formulas

$$(6.8) \quad \begin{array}{l} b_1 = -a_1 a_2, \quad b_2 = a_2, \ b_{2p} b_{2p+1} = -a_{p+2}, \quad b_{2p+1} = -b_{2p+2}, \\ \beta_1 = \alpha_1 + \alpha_2, \ \beta_2 = \alpha_2, \ \beta_{2p} + \beta_{2p+1} = \alpha_{p+2}, \ \beta_{2p+1} = \beta_{2p+2}, \end{array} \quad (p = 1, 2, 3, \cdots),$$

so that (6.7) becomes

$$(6.9) \quad 1 + a_1 z^{\alpha_1} - \frac{a_1 a_2 z^{\alpha_1 + \alpha_2}}{1 + a_2 z^{\alpha_2}} + \frac{a_3 z^{\alpha_3}}{1} + \frac{a_4 z^{\alpha_4}}{1} + \cdots.$$

By (6.5), the equations (6.8) can be solved for the β_p, and give (6.6). One may now readily verify that the approximants of (6.9) are the 1-st, 2-nd, 3-rd, $\cdots$ approximants of (1.1). It then follows immediately that the power series for (6.2), in which the b_p and β_p are given by (6.6), is the same as the power series for (1.1), and the theorem is proved.

Example. Theorem 6.1 may be applied to the C-fraction of Ramanujan ([3], p. 214), and gives the identity

$$(6.10)\quad 1+\frac{z}{1}+\frac{z^2}{1}+\frac{z^3}{1}+\frac{z^4}{1}+\cdots$$
$$=1+z-\frac{z^3}{1}+\frac{z^2}{1}-\frac{z}{1}+\frac{z}{1}-\frac{z^3}{1}+\frac{z^3}{1}-\frac{z^2}{1}+\frac{z^2}{1}-\frac{z^4}{1}+\frac{z^4}{1}-\cdots.$$

THEOREM 6.2. *If the exponents α_p of the C-fraction* (1.1) *satisfy the inequalities*

$$(6.11)\quad \alpha_1+\alpha_3+\cdots+\alpha_{2p+1}>\alpha_2+\alpha_4+\cdots+\alpha_{2p}>\alpha_1+\alpha_3+\cdots+\alpha_{2p-1},\quad (p=1,2,3,\cdots),$$

then the coefficients d_p and exponents δ_p in (6.3) *are given by the formulas*

$$(6.12)\quad \begin{aligned} &d_1=-d_2=-a_1,\quad d_{4p-1}=-d_{4p}=-\frac{a_2a_4\cdots a_{2p}}{a_1a_3\cdots a_{2p-1}},\\ &d_{4p+1}=-d_{4p+2}=-\frac{a_1a_3\cdots a_{2p+1}}{a_2a_4\cdots a_{2p}},\\ &\delta_1=\delta_2=\alpha_1,\ \delta_{4p-1}=\delta_{4p}=(\alpha_2+\alpha_4+\cdots+\alpha_{2p})-(\alpha_1+\alpha_3+\cdots+\alpha_{2p-1}),\\ &\delta_{4p+1}=\delta_{4p+2}=(\alpha_1+\alpha_3+\cdots+\alpha_{2p+1})-(\alpha_2+\alpha_4+\cdots+\alpha_{2p}),\\ &\qquad (p=1,2,3,\cdots). \end{aligned}$$

Proof. The odd part of (6.3), in which the d_p and δ_p are given by (6.12), is precisely (1.1). Hence (6.3) and (1.1) must have one and the same power series.

Example. If we apply Theorem 6.2 to the C-fraction of Ramanujan, we obtain the identity

$$(6.13)\quad 1+\frac{z}{1}+\frac{z^2}{1}+\frac{z^3}{1}+\cdots$$
$$=\frac{1}{1}-\frac{z}{1}+\frac{z}{1}-\frac{z}{1}+\frac{z}{1}-\frac{z^2}{1}+\frac{z^2}{1}-\frac{z^2}{1}+\frac{z^2}{1}-\frac{z^3}{1}+\frac{z^3}{1}-\frac{z^3}{1}+\frac{z^3}{1}-\cdots.$$

The approximants of (6.2) and (6.3) are of course rational approximants for $P(z)$, and are easily seen to be Padé approximants in case the power series $P^*(z)$ and $Q(z)$ of (6.1) are regular. Under the conditions of Theorems 6.1 and 6.2 it may be verified by means of Theorem 4.2 that $P^*(z)$ and $Q(z)$, respectively, are α-regular. Hence we have the following theorem:

THEOREM 6.3. *Under the conditions of Theorems* 6.1 *and* 6.2, *the approximants of* (6.2) *and* (6.3), *respectively, are all Padé approximants for* $P(z)$.

Example. In Fig. 3 are shown blocks in the Padé table for the Ramanujan C-fraction, which are occupied by the approximants of the latter and of the continued fractions in the right-hand members of (6.10) and (6.13). These blocks are designated by $0, 1, 2, \cdots$; $0^*, 1^*, 2^*, \cdots$; and ${}^*0, {}^*1, {}^*2, \cdots$,

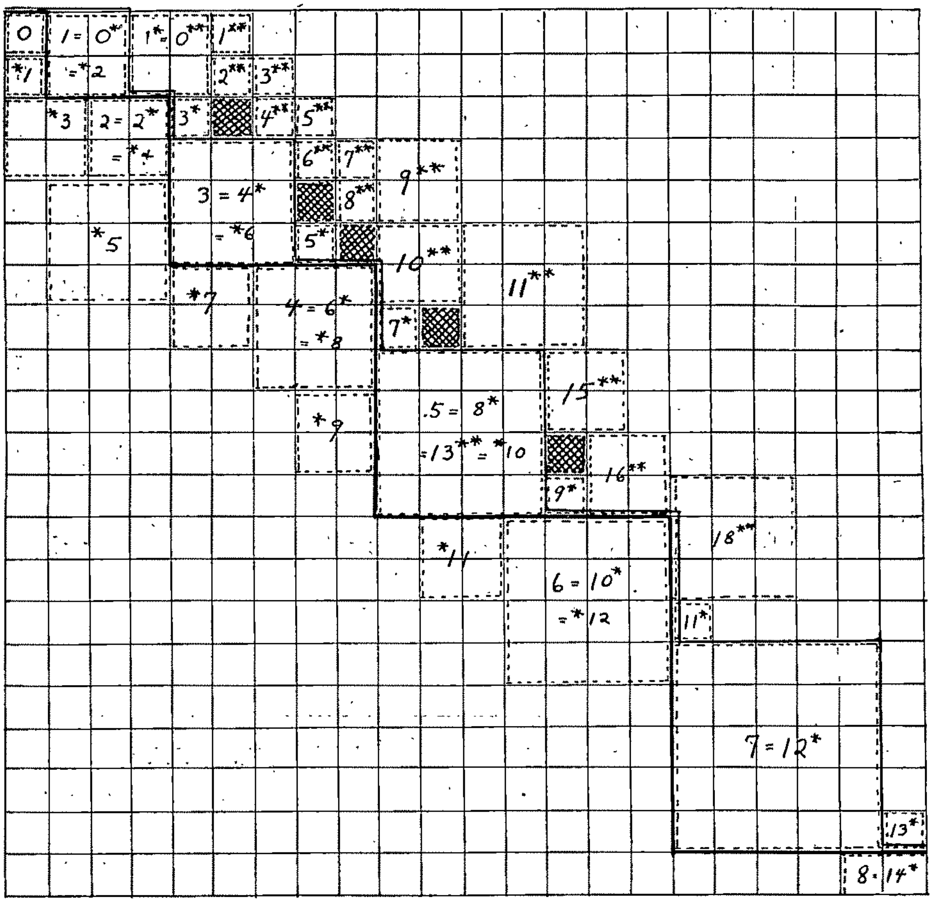


Fig. 3

respectively. The blocks designated by $0^{**}, 1^{**}, 2^{**}, \cdots$ are occupied by certain of the approximants of

$$(6.14)\quad 1+z-z^3+\frac{z^5}{1}-\frac{z}{1}+\frac{2z}{1}-\frac{3z/2}{1}-\frac{z/6}{1}-\frac{4z/3}{1}+\frac{3z}{1}-\frac{\frac{1}{2}z}{1}-\frac{\frac{1}{2}z}{1}$$
$$-\frac{z}{1}+\frac{2z^2}{1}-\frac{z^3}{1}+\frac{z}{1}+\frac{z}{1}-\frac{z}{1}+\frac{z}{1}-\frac{z^3}{1}-\frac{z}{1}+\frac{z^3}{1}+\cdots,$$

which is also equal to the C-fraction of Ramanujan. This can be obtained by

means of the algorithm of 2, if we start with the power series for the Ramanujan C-fraction, namely,

$$P(z)=1+z-z^3+z^5+z^6-z^7-2z^8+2z^{10}+2z^{11}-z^{12}-3z^{13}-z^{14}$$
$$+3z^{15}+3z^{16}-2z^{17}-5z^{18}-z^{19}+6z^{20}+5z^{21}-3z^{22}-8z^{23}$$
$$-2z^{24}+8z^{25}+7z^{26}-5z^{27}-12z^{28}-2z^{29}+13z^{30}+\cdots.$$

The approximants of (6.14) are not all Padé approximants, e. g., the 12-th is not. The squares which are shaded in Fig. 3 are not occupied by approximants of any of these continued fractions. This example shows that when the process by which (6.2) was obtained is repeated, the new continued fractions may have approximants which are not all Padé approximants although (1.1) is regular.

In case the $\alpha_p=1$, Stieltjes gave formulas for computing (6.2), (6.3), (6.4), which hold under certain conditions. The formulas we have given in Theorems 6.1 and 6.2 do not apply when the $\alpha_p=1$. The following theorems may be regarded as generalizations of the results of Stieltjes.

THEOREM 6.4. *If the exponents α_p of the C-fraction* (1.1) *satisfy the condition*

$$(6.15)\quad \alpha_3+\alpha_5+\cdots+\alpha_{2p+1}=\alpha_2+\alpha_4+\cdots+\alpha_{2p},\quad (p=1,2,3,\cdots),$$

then the coefficients b_p and exponents β_p in (6.2) *are given by the formulas*

$$b_1=-a_1a_2,\quad b_2=a_2+a_3,\quad b_{2p+1}=a_{2p+3}+a_{2p+2}-b_{2p+2},$$

$$(6.16)\quad b_{2p+2}=a_{2p+3}+\cfrac{a_{2p+2}}{1+\dfrac{a_{2p+1}}{a_{2p}}+\dfrac{a_{2p+1}a_{2p-1}}{a_{2p}a_{2p-2}}+\cdots+\dfrac{a_{2p+1}a_{2p-1}\cdots a_3}{a_{2p}a_{2p-2}\cdots a_2}},$$
$$(p=1,2,3,\cdots);$$

$$\beta_1=\alpha_1+\alpha_2,\quad \beta_2=\alpha_2=\alpha_3,\quad \beta_{2p-1}=\beta_{2p}=\alpha_{2p}=\alpha_{2p+1},$$
$$(p=2,3,4,\cdots),$$

provided the values of the a_p are such that (6.16) *can be solved for the $b_p\neq 0$.*

Theorem 6.4 may be proved by comparing the odd part of (1.1) with the even part of (6.2).

THEOREM 6.5. *If the exponents α_p of* (1.1) *satisfy the condition*

$$(6.17)\quad \alpha_2+\alpha_4+\cdots+\alpha_{2p}=\alpha_1+\alpha_3+\cdots+\alpha_{2p-1},\quad (p=1,2,3,\cdots),$$

then the d_p and δ_p in (6.3) *are given by the formulas*

$$
\begin{aligned}
& d_1 = -a_1, \quad d_2 = a_1 + a_2, \\
(6.18) \qquad & d_{2p} d_{2p+1} = a_{2p} a_{2p+1}, \quad d_{2p+1} + d_{2p+2} = a_{2p+1} + a_{2p+2}, \\
& \delta_1 = \delta_2 = \alpha_1 = \alpha_2, \quad \delta_{2p+1} = \delta_{2p+2} = \alpha_{2p+1} = \alpha_{2p+2}, \quad (p = 1, 2, 3, \cdots),
\end{aligned}
$$

provided the a_p are such that (6.18) can be solved for the $d_p \neq 0$.

To prove Theorem 6.5, we compare the even part of (1.1) with the odd part of (6.3).

If we change the notation by substituting

$$a_1 = 1/k_1, \quad a_p = 1/k_{p-1}k_p, \quad d_1 = 1/f_1, \quad d_p = 1/f_{p-1}f_p, \quad (p = 2, 3, 4, \cdots),$$

we find that formulas (6.18) may be written in the form

$$
\begin{aligned}
& f_1 = -k_1, \quad f_2 = -k_2/(1 + k_2), \\
(6.19) \qquad & f_{2p+1} = -k_{2p+1}(1 + k_2 + k_4 + \cdots + k_{2p})^2, \quad (p = 1, 2, 3, \cdots), \\
& f_{2p} = \frac{-k_{2p}}{(1 + k_2 + k_4 + \cdots + k_{2p-2})(1 + k_2 + k_4 + \cdots + k_{2p})}, \\
& \qquad (p = 2, 3, 4, \cdots).
\end{aligned}
$$

It must, of course, be assumed that $1 + k_2 + k_4 + \cdots + k_{2p} \neq 0$, $(p = 1, 2, 3, \cdots)$. These formulas are exactly the same as those found by Stieltjes in the case where the $\alpha_p = 1$.

THEOREM 6.6. *If the exponents α_p of (1.1) satisfy (6.15), then the coefficients e_p and exponents ϵ_p of (6.4) are given by the formulas*

$$
\begin{aligned}
(6.20) \qquad & e_1 = -a_1 a_2, \; e_2 = a_2 + a_3, \; e_{2p} e_{2p+1} = a_{2p+1} a_{2p+2}, \; e_{2p+1} + e_{2p+2} = a_{2p+2} + a_{2p+3}, \\
& \epsilon_1 = \alpha_1 + \alpha_2, \; \epsilon_2 = \alpha_2 = \alpha_3, \; \epsilon_{2p+1} = \epsilon_{2p+2} = \alpha_{2p+2} = \alpha_{2p+3}, \\
& \qquad (p = 1, 2, 3, \cdots),
\end{aligned}
$$

provided $a_1 = -C$. If $a_1 \neq -C$, the formulas relating the a_p, α_p, e_p, and ϵ_p are

$$
\begin{aligned}
(6.21) \qquad & e_1 = a_1 + C, \quad e_{2p-1} e_{2p} = a_{2p-1} a_{2p}, \quad e_{2p} + e_{2p+1} = a_{2p} + a_{2p+1}, \\
& \epsilon_1 = \alpha_1, \quad \epsilon_{2p} = \epsilon_{2p+1} = \alpha_{2p} = \alpha_{2p+1}, \qquad (p = 1, 2, 3, \cdots).
\end{aligned}
$$

The values of the a_p must be such that conditions (6.20) and (6.21) are solvable for the $e_p \neq 0$.

Proof. The even part of (6.4), in which the e_p and ϵ_p are given by (6.20), is the same as the odd part of (1.1). Therefore (6.4) and (1.1) must have the same power series. Similarly, the odd part of (6.4), in which the e_p and ϵ_p are given by (6.21), is precisely the odd part of (1.1). Again the power series for (6.4) and (1.1) must be the same.

Formulas (6.21) are equivalent to those found by Stieltjes for the case where the $\alpha_p = 1$. For, if we let

$$a_1 = 1/k_1, \quad a_p = 1/k_{p-1}k_p, \quad e_1 = 1/h_1, \quad e_p = 1/h_{p-1}h_p, \quad (p = 2, 3, 4, \cdots),$$

then we find that formulas (6.21) become

$$(6.22) \quad \begin{gathered} h_1 = k_1/(1 + Ck_1), \quad h_{2p} = k_{2p}(1 + C(k_1 + k_3 + \cdots + k_{2p-1}))^2, \\ h_{2p+1} = \frac{k_{2p+1}}{(1 + C(k_1 + k_3 + \cdots + k_{2p-1}))(1 + C(k_1 + k_3 + \cdots + k_{2p+1}))}, \\ (p = 1, 2, 3, \cdots). \end{gathered}$$

One may readily verify the following theorem by means of Theorem 4.2.

THEOREM 6.7. *Under the conditions of Theorems* 6.4, 6.5, *and* 6.6, *the approximants of* (6.2), (6.3), *and* (6.4), *respectively, are all Padé approximants for* $P(z)$.

By analogous devices, one can obtain the continued fractions (6.2), (6.3), and (6.4) in other cases, e.g., where the α_p satisfy (6.11) for certain values of p and (6.15) for the remaining values. To do this, one must form, from the odd and even parts of (1.1), continued fractions which have all but a finite number of the approximants of (1.1), as, for instance, the continued fraction (6.9).

NORTHWESTERN UNIVERSITY,
EVANSTON, ILLINOIS.

BIBLIOGRAPHY.

1. Walter Leighton and W. T. Scott, "A general continued fraction expansion," *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 596-605.

2. O. Perron, *Die Lehre von den Kettenbrüchen*, Leipzig, 1929.

3. Ramanujan, *Collected Papers of Srinivasa Ramanujan*, edited by G. H. Hardy, P. V. Seshu Aiyar, and B. M. Wilson, Cambridge, 1927.

4. W. T. Scott and H. S. Wall, "Continued fraction expansions for arbitrary power series," *Annals of Mathematics*, vol. 41 (1940), pp. 328-349.

5. T. J. Stieltjes, *Oeuvres*, vol. 2, Groningen, 1918.

6. H. S. Wall, "Note on the expansion of a power series into a continued fraction," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 97-105.

THE BERNSTEIN POLYNOMIALS FOR DISCONTINUOUS FUNCTIONS.*

By Fritz Herzog and J. D. Hill.

1. **Introduction.** A classical theorem of Weierstrass asserts that every continuous function defined on a finite closed interval can be uniformly approximated by polynomials. There are several well known proofs of this theorem in the literature; among these is the very elegant demonstration of S. Bernstein,[1] who showed that if $f(x)$ is continuous in $[0,1]$[2] then the sequence of polynomials

$$B_n(f;x) \equiv \sum_{k=0}^{n} f(k/n) T_{nk}(x) \qquad (n=1,2,3,\cdots), \tag{1.1}$$

where

$$T_{nk}(x) \equiv C_{n,k} x^k (1-x)^{n-k}, \tag{1.2}$$

converges uniformly to $f(x)$ in $[0,1]$.

For later reference it is convenient at this point to observe the following obvious facts.

(1.3) $a \leqq f(k/n) \leqq A$ for all k/n in $[0,1]$ implies
$a \leqq B_n(f;x) \leqq A$ for all x in $[0,1]$.

$$B_n(f;0) = f(0); \qquad B_n(f;1) = f(1). \tag{1.4}$$

The elegance of Bernstein's theorem lies in the facts that the approximating polynomials are given explicitly and that they depend only on the values of the function $f(x)$ for rational values of x. This latter fact leads us to introduce at once the following definitions.

(1.5) *Definition.* By a *skeleton* we shall mean a function $f(r)$ defined only for rational values of r in $[0,1]$.

On the other hand, in this paper we shall hereafter reserve the term *function* to denote an $f(x)$ defined for all x in $[0,1]$.

(1.6) *Definition.* By the *class* $\mathfrak{S}$ we shall mean the class of all skeletons $f(r)$ for which the limits

* Received September 4, 1945.

[1] S. Bernstein, *Communications de la Société Mathématique de Kharkov*, (2), vol. 13 (1912), pp. 1-2.

[2] By $[a,b]$ or (a,b) we mean the interval $a \leqq x \leqq b$ or $a < x < b$, respectively.

$$(1.7) \qquad f_L(x) \equiv \lim_{r\to x-} f(r) \qquad (0 < x \leqq 1),$$

$$(1.8) \qquad f_R(x) \equiv \lim_{r\to x+} f(r) \qquad (0 \leqq x < 1),$$

exist for the values of x indicated.

(1.9) *Definition.* For each $f \in \mathfrak{S}$ we shall understand by the *normalized extension* of $f(r)$ the function $f_N(x)$ defined by the use of (1.7) and (1.8) as follows.

$$(1.10) \qquad f_N(x) \equiv \tfrac{1}{2}[f_L(x) + f_R(x)] \qquad (0 < x < 1),$$

$$(1.11) \qquad f_N(0) \equiv f(0); \qquad f_N(1) \equiv f(1).$$[3]

(1.12) *Definition.* Two arbitrary skeletons $f(r)$ and $g(r)$ are called *equivalent*, and we write $f(r) \sim g(r)$, if for each $\epsilon > 0$ the inequality $|f(r) - g(r)| \geqq \epsilon$ holds for at most a finite number of values of r.

We are now able to state precisely the object of this paper. We propose to investigate the behavior of the polynomials (1.1), which we shall call the *Bernstein polynomials* of f, for the case of a bounded skeleton f.

In **2** we collect in the form of lemmas certain results which are used throughout. In **3** we establish some useful lemmas concerning skeletons of the class $\mathfrak{S}$ and their normalized extensions.

The behavior of the Bernstein polynomials of a skeleton f, under the single restriction that f be bounded, is discussed in **4**. Here we obtain as a principal result Theorem (4.3), which gives lower and upper bounds, respectively, for $\underline{\lim}_n B_n(f;x)$ and $\overline{\lim}_n B_n(f;x)$ for each x. In **5** we derive the following results concerning skeletons f of the class $\mathfrak{S}$. In the first place we show that the Bernstein polynomials of such a skeleton converge in $[0,1]$ to its normalized extension (Theorem (5.1)); and that this convergence is uniform in any closed subinterval in which the normalized extension is continuous (Theorem (5.5)). In the second place we show that if f and g are two skeletons in $\mathfrak{S}$ then $B_n(f;x)$ and $B_n(g;x)$ converge to the same limit function if and only if f and g are equivalent (Theorem (5.6)).

In **6** we derive sufficient conditions for the convergence of the Bernstein polynomials of an arbitrary skeleton f in terms of the order of magnitude of the sum $\sum_{k=0}^{n} |f(k/n) - g(k/n)|$, where g is a skeleton in $\mathfrak{S}$. A necessary condition for convergence in terms of the order of magnitude of $\sum_{k=0}^{n} [f(k/n) - g(k/n)]$ is also given.

---

[3] This somewhat unnatural definition at the endpoints is justified by its convenience See (1.4).

Finally in 7 we present some examples to illustrate the main features of the theory.

2. Preliminary lemmas. The majority of our results depend upon one or more of the following lemmas. For the definition of $T_{nk}(x)$ we refer the reader to (1.2).

(2.1) LEMMA. *For $n=1,2,3,\cdots$; $k=0,1,2,\cdots,n$ and $0<x<1$, we have $T_{nk}(x)<K[nx(1-x)]^{-\frac{1}{2}}$, where K is an absolute constant.*

Proof. We shall show that $n^{\frac{1}{2}}C_{n,k}x^{k+\frac{1}{2}}(1-x)^{n-k+\frac{1}{2}}$ is bounded for all values of n, k and x. As a function of x this expression assumes its maximum value for $x=(k+\frac{1}{2})/(n+1)$. This maximum value may be written in the form

$$\frac{n^{\frac{1}{2}}n!}{(n+1)^{n+1}}\ \frac{(k+\frac{1}{2})^{k+\frac{1}{2}}}{k!}\ \frac{(n-k+\frac{1}{2})^{n-k+\frac{1}{2}}}{(n-k)!}. \tag{2.2}$$

The second and third factors in (2.2) have the form $A_m=(m+\frac{1}{2})^{m+\frac{1}{2}}/m!$. By the use of Stirling's formula we find that $A_m<K_1e^m$ for $m\geqq 0$, and that the first factor in (2.2) is less than K_2e^{-n}, where K_1 and K_2 are certain constants. Combining these estimates we find that the expression in (2.2) is less than a constant K.

(2.3) LEMMA. *For $n=1,2,3,\cdots$ and $0\leqq x\leqq 1$ we have $\sum_k^* T_{nk}(x)<1/(4n^{\frac{1}{2}})$, where the sum is taken over all k for which $|k-nx|>n^{3/4}$.*

This is a known result.[4]

(2.4) LEMMA. *For $n=1,2,3,\cdots$ and $0<x<1$ we have*

$$\lim_{n\to\infty}\sum_{0\leqq k<nx}T_{nk}(x)=\tfrac{1}{2}=\lim_{n\to\infty}\sum_{nx<k\leqq n}T_{nk}(x).$$

This follows at once from the central limit theorem of probability, namely, that

$$\lim_{n\to\infty}\sum_k T_{nk}(x)=(2\pi)^{-\frac{1}{2}}\int_{-\infty}^{t}e^{-u^2/2}du,$$

where the sum is taken over those values of k for which $0\leqq k<nx+t[nx(1-x)]^{\frac{1}{2}}$. Merely set $t=0$.

(2.5) LEMMA. *For $n=1,2,3,\cdots$ and $0<x<1$ we have*

$$\lim_{n\to\infty}\sum_k{}' T_{nk}(x)=\tfrac{1}{2}=\lim_{n\to\infty}\sum_k{}'' T_{nk}(x),$$

---

[4] G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. 1, Berlin (1925), p. 66, no. 145.

where Σ_k' is taken over all k for which $nx - n^{3/4} \leqq k < nx$ and Σ_k'' over all k for which $nx < k \leqq nx + n^{3/4}$.

This is a direct consequence of Lemmas (2.3) and (2.4).

(2.6) LEMMA. *If $F(x)$ is bounded in $[0,1]$ and continuous at a point x_0 in $(0,1)$ then the sequence $\{B_n(F;x_0)\}$ converges to $F(x_0)$. Moreover if $F(x)$ is continuous at every point*[5] *of $[a,b]$ where $0 \leqq a < b \leqq 1$, then the Bernstein polynomials $B_n(F;x)$ converge uniformly to $F(x)$ in $[a,b]$.*

For the case in which $a = 0$ and $b = 1$ this is precisely Bernstein's original theorem (see **1**). In the general case appropriate modifications of the proof given by Pólya and Szegö[6] suffice to establish the lemma as stated.

3. The class $\mathfrak{S}$. In (1.5) we introduced the concept of a skeleton and in (1.6) we defined the class $\mathfrak{S}$ of skeletons. We now proceed to establish several results concerning skeletons of the class $\mathfrak{S}$. The first of these is the following lemma.

(3.1) LEMMA. *If $f(r) \in \mathfrak{S}$ then $f_L(x)$ and $f_R(x)$ defined by (1.7) and (1.8) are at most simply discontinuous functions for $0 < x \leqq 1$ and $0 \leqq x < 1$, respectively, and*

$$(3.2) \qquad f_L(x-) = f_R(x-) = f_L(x) \qquad (0 < x \leqq 1),$$

$$(3.3) \qquad f_L(x+) = f_R(x+) = f_R(x) \qquad (0 \leqq x < 1).$$

Proof. Let x_0 be a fixed point in the interval $0 < x \leqq 1$. Equation (3.2) is equivalent to the assertion that for an arbitrary sequence $\{x_n\}$ converging to x_0 from the left, we have

$$(3.4) \qquad \lim_{n\to\infty} f_L(x_n) = \lim_{n\to\infty} f_R(x_n) = f_L(x_0).$$

To establish (3.4) we associate with each x_n two rational numbers r_n' and r_n'', such that $x_n - 1/n < r_n' < x_n < r_n'' < x_0$ and such that

$$(3.5) \qquad |f(r_n') - f_L(x_n)| < 1/n; \quad |f(r_n'') - f_R(x_n)| < 1/n.$$

We now make use of the following identities.

$$(3.6) \qquad f_L(x_n) = f_L(x_0) + [f(r_n') - f_L(x_0)] - [f(r_n') - f_L(x_n)],$$

$$(3.7) \qquad f_R(x_n) = f_L(x_0) + [f(r_n'') - f_L(x_0)] - [f(r_n'') - f_R(x_n)].$$

---

[5] By this we mean that if $a > 0$ then $F(x)$ is to be continuous on the left at a as well as on the right; an analogous statement holds with respect to b if $b < 1$.

[6] G. Pólya and G. Szegö, *loc. cit.*, p. 66, no. 146.

Since $r_n' \to x_0 -$ and $r_n'' \to x_0 -$ the differences in the first brackets on the right sides of (3.6) and (3.7) approach zero, while the second brackets do so because of (3.5). Consequently (3.6) and (3.7) imply (3.4). In a similar fashion we may establish (3.3).

The next lemma, which is a simple consequence of the preceding, involves the notion of the normalized extension (see (1.9)) of a skeleton $f \in \mathfrak{S}$.

(3.8) LEMMA. *The normalized extension $f_N(x)$ of a skeleton $f(r) \in \mathfrak{S}$ is at most a simply discontinuous function in* $[0,1]$. *Moreover, the following relations hold.*

$$f_N(x-) = f_L(x) \qquad (0 < x \leqq 1), \tag{3.9}$$

$$f_N(x+) = f_R(x) \qquad (0 \leqq x < 1), \tag{3.10}$$

$$f_N(x) = \tfrac{1}{2}[f_N(x-) + f_N(x+)] \qquad (0 < x < 1). \tag{3.11}$$

Proof. Let x_0 be any point in the interval $0 < x \leqq 1$. By Lemma (3.1) the limits $f_L(x_0 -)$ and $f_R(x_0 -)$ exist, so that $f_N(x_0 -)$ exists by (1.10). Equation (3.9) then follows at once from (1.10) and (3.2). Similarly, for an x_0 in the interval $0 \leqq x < 1$, $f_N(x_0 +)$ exists and (3.10) holds. Equation (3.11) now follows directly from (1.10), (3.9) and (3.10).

For a skeleton $f(r) \in \mathfrak{S}$ the values of $f_N(r)$ and $f(r)$ are not necessarily equal. For example, if $f(r) \equiv 1/m$ when $r = h/m$ in its lowest terms, then $f_N(x) = 0$ in $(0,1)$, and thus $f_N(r)$ is actually different from $f(r)$ for all values of r in $(0,1)$. The following lemma, however, establishes the fact that $f(r)$ and $f_N(r)$ are always equivalent in the sense of (1.12).

(3.12) LEMMA. *If $f(r)$ is a skeleton of the class $\mathfrak{S}$ and if $f_N(x)$ is its normalized extension, then $f_N(r) \sim f(r)$.*

Proof. According to Definition (1.12) we have to show that for each $\epsilon > 0$ the inequality $| f_N(r) - f(r) | \geqq \epsilon$ is satisfied by at most a finite number of values of r. Suppose, on the contrary, that for a certain $\epsilon_0 > 0$ the inequality $| f_N(r) - f(r) | \geqq \epsilon_0$ holds for infinitely many r. Among these values of r there is a sequence $\{r_n\}$ such that either $r_n \to x_0 -$ or $r_n \to x_0 +$, where x_0 is some point in $[0,1]$. We assume for definiteness that $r_n \to x_0 -$ so that $0 < x_0 \leqq 1$. By (3.9) we have $\lim_n f_N(r_n) = f_L(x_0)$, and by (1.7) we have $\lim_n f(r_n) = f_L(x_0)$. Consequently we find that $\lim_n [f_N(r_n) - f(r_n)] = 0$, which contradicts the assumption that $| f_N(r_n) - f(r_n) | \geqq \epsilon_0$ for all n. A similar argument holds if $r_n \to x_0 +$.

The connection between equivalent skeletons in $\mathfrak{S}$ and their normalized extensions is shown by the following result.

(3.13) LEMMA. *Two skeletons in $\mathfrak{S}$ are equivalent if and only if their normalized extensions are identical in* $(0,1)$.

Proof. Let $f(r)$ and $g(r)$ be two skeletons in $\mathfrak{S}$ and let $f_N(x)$ and $g_N(x)$ be their normalized extensions. Then if we set $h(r) = f(r) - g(r)$ it follows from (1.10) that $h_N(x) = f_N(x) - g_N(x)$ in $(0,1)$. Now if $h_N(x) = 0$ in $(0,1)$ then $h(r) \sim 0$ by Lemma (3.12) and hence $f(r) \sim g(r)$. Conversely, if $f(r) \sim g(r)$ then $h(r) \sim 0$, from which it follows at once that $h_N(x) = 0$ in $(0,1)$.

4. The Bernstein polynomials for bounded skeletons. For any skeleton $f(r)$, bounded or not, we now define the following four auxiliary functions.

(4.1) $$\underline{f}_L(x) \equiv \varliminf_{r\to x-} f(r); \quad \overline{f}_L(x) \equiv \varlimsup_{r\to x-} f(r) \qquad (0 < x \leqq 1),$$

(4.2) $$\underline{f}_R(x) \equiv \varliminf_{r\to x+} f(r); \quad \overline{f}_R(x) \equiv \varlimsup_{r\to x+} f(r) \qquad (0 \leqq x < 1).$$

It can be easily shown that the finiteness of these four limits for the values of x indicated is sufficient as well as necessary for the boundedness of $f(r)$.

We recall that the Bernstein polynomials $B_n(f;x)$ for an arbitrary skeleton $f(r)$ were defined in (1.1). We come now to our first main result which gives the best possible restriction on the behavior of the sequence $\{B_n(f;x)\}$ in $(0,1)$, as f ranges over the class of bounded skeletons. On account of (1.4) there is no convergence problem for $x = 0$ or 1.

(4.3) THEOREM. *For an arbitrary bounded skeleton $f(r)$ the following inequalities hold for $0 < x < 1$.*

(4.4) $$\varliminf_{n\to\infty} B_n(f;x) \geqq \tfrac{1}{2}[\underline{f}_L(x) + \underline{f}_R(x)],$$

(4.5) $$\varlimsup_{n\to\infty} B_n(f;x) \leqq \tfrac{1}{2}[\overline{f}_L(x) + \overline{f}_R(x)].$$

Proof. We shall give the proof of (4.5). The proof of (4.4) will then follow by applying (4.5) to $-f$. Let x_0 be a fixed point in $(0,1)$. We shall write (1.1) in the form

(4.6) $$B_n(f;x_0) = \sum_k{}' f(k/n)T_{nk}(x_0) + \sum_k{}'' f(k/n)T_{nk}(x_0) + \sum_k{}^* f(k/n)T_{nk}(x_0) + X,$$

where here and in what follows the symbols $\sum_k{}'$, $\sum_k{}''$, $\sum_k{}^*$ denote summation

over those values of k for which $nx_0 - n^{3/4} \leqq k < nx_0$, $nx_0 < k \leqq nx_0 + n^{3/4}$, $|k - nx_0| > n^{3/4}$, respectively. The term X corresponds to $k = nx_0$ in case nx_0 is an integer and is equal to 0 otherwise. In either case by Lemma (2.1) we have $X = o(1)$ as $n \to \infty$.

Now let λ_n' and λ_n'' denote the least upper bounds of $f(r)$ for $x_0 - n^{-1/4} \leqq r < x_0$ and $x_0 < r \leqq x_0 + n^{-1/4}$, respectively. By (4.1) and (4.2) we have

$$\text{(4.7)} \qquad \lim_{n\to\infty} \lambda_n' = \overline{f}_L(x_0); \quad \lim_{n\to\infty} \lambda_n'' = \overline{f}_R(x_0).$$

Furthermore, from the definitions of $\sum_k'$ and $\sum_k''$, we have

$$\text{(4.8)} \qquad \sum_k{}' f(k/n) T_{nk}(x_0) \leqq \lambda_n' \sum_k{}' T_{nk}(x_0),$$

$$\text{(4.9)} \qquad \sum_k{}'' f(k/n) T_{nk}(x_0) \leqq \lambda_n'' \sum_k{}'' T_{nk}(x_0).$$

By (4.7) and Lemma (2.5) the right sides of (4.8) and (4.9) approach $\frac{1}{2}\overline{f}_L(x_0)$ and $\frac{1}{2}\overline{f}_R(x_0)$, respectively, as $n \to \infty$. Finally, if M is an upper bound for $f(r)$ then

$$\text{(4.10)} \qquad \sum_k{}^* f(k/n) T_{nk}(x_0) \leqq M \sum_k{}^* T_{nk}(x_0).$$

By Lemma (2.3) the right side of (4.10) tends to 0 as $n \to \infty$. The preceding estimates taken in conjunction with (4.6) clearly establish (4.5).

(4.11) Corollary. *For an arbitrary bounded skeleton $f(r)$ the following inequalities*[7] *hold for $0 < x < 1$.*

$$\varliminf_{r\to x} f(r) \leqq \varliminf_{n\to\infty} B_n(f;x) \leqq \varlimsup_{n\to\infty} B_n(f;x) \leqq \varlimsup_{r\to x} f(r).$$

In 7 we shall give examples (see (7.1) and (7.3)) which show that for a bounded skeleton $f(r)$ no more than (4.4) and (4.5) can be said in general. Indeed, by a combination of examples of the types (7.1) and (7.3) it is possible to establish the following fact: If x_0 is any point in $(0,1)$ then there exists a skeleton $f(r)$ for which the six quantities $\underline{f}_L(x_0)$, $\overline{f}_L(x_0)$, $\underline{f}_R(x_0)$, $\overline{f}_R(x_0)$, $\varliminf_n B_n(f;x_0)$ and $\varlimsup_n B_n(f;x_0)$ assume any six given values a', A', a'', A'', λ and Λ, respectively, for which $\frac{1}{2}(a' + a'') \leqq \lambda \leqq \Lambda \leqq \frac{1}{2}(A' + A'')$, $a' \leqq A'$ and $a'' \leqq A''$.

5. The Bernstein polynomials for skeletons in $\mathfrak{S}$. In this section we shall be concerned for the most part with skeletons f of the class $\mathfrak{S}$. There

---

[7] By "$\varliminf_{r\to x} f(r)$" we mean the limit as δ approaches 0 of the greatest lower bound of $f(r)$ for $0 < |r - x| < \delta$; and analogously for the upper limit.

are several questions that naturally arise in connection with such skeletons. The first of these is the following: Does the sequence $\{B_n(f;x)\}$ converge and, if so, what is the limit function? This question is completely answered in the next theorem.

(5.1) THEOREM. *If $f(r)$ is a skeleton of class $\mathfrak{S}$ the sequence of its Bernstein polynomials converges for $0 \leqq x \leqq 1$ to $f_N(x)$, the normalized extension of $f(r)$.*

Proof. By (4.1) and (1.7), $\underline{f}_L(x) = \bar{f}_L(x) = f_L(x)$ and, by (4.2) and (1.8), $\underline{f}_R(x) = \bar{f}_R(x) = f_R(x)$ for $0 < x < 1$. Hence the right sides of (4.4) and (4.5) are equal and their common value, by (1.10), is $f_N(x)$. Thus from Theorem (4.3) we conclude that $\lim_n B_n(f;x)$ exists for $0 < x < 1$ and equals $f_N(x)$. At $x = 0$ and $x = 1$ the result is clear from (1.4) and (1.11).

It is of interest to observe in passing that the proof of the preceding theorem serves to establish the following result, which is local in character and does not require that f belong to $\mathfrak{S}$. If $f(r)$ is a bounded skeleton and if at a point x_0 in $(0,1)$ the limits in (1.7) and (1.8) exist then we have $\lim_n B_n(f;x_0) = \frac{1}{2}[f_L(x_0) + f_R(x_0)]$.

Before investigating the question of when the convergence assured by Theorem (5.1) is uniform, we find it convenient to establish the following lemma.

(5.2) LEMMA. *Let $g(r)$ be a skeleton satisfying the following conditions:* (i) *$g(r)$ is bounded;* (ii) *$g_R(0) = g(0) = 0$ and $g_L(1) = g(1) = 0$;* (iii) *for each fixed $\epsilon > 0$, $\sum_k^{(1)} |g(k/n)| = o(n^{\frac{1}{2}})$ as $n \to \infty$, where the symbol $\sum_k^{(1)}$ denotes summation over those values of k for which $|g(k/n)| \geqq \epsilon$. Then the sequence $\{B_n(g;x)\}$ converges uniformly to zero in $[0,1]$.*

Proof. Let $\epsilon > 0$ be fixed arbitrarily. By condition (ii) there exists a δ $(0 < \delta < 1/4)$ such that $|g(r)| < \epsilon$ for $0 \leqq r \leqq 2\delta$ and for $1 - 2\delta \leqq r \leqq 1$. For each $n = 1, 2, 3, \cdots$ we write the Bernstein polynomial of $g(r)$ in the form

$$(5.3) \qquad B_n(g;x) = \sum_k^{(1)} g(k/n) T_{nk}(x) + \sum_k^{(2)} g(k/n) T_{nk}(x),$$

where $\sum_k^{(1)}$ has already been defined in (iii) above and $\sum_k^{(2)}$ is extended to those values of k for which $|g(k/n)| < \epsilon$. It is obvious that the second term on

the right in (5.3) does not exceed ϵ in absolute value for $0 \leqq x \leqq 1$ and $n = 1, 2, 3, \cdots$. In order to estimate the first term on the right in (5.3) we consider first the case in which $\delta \leqq x \leqq 1 - \delta$. In this case, by applying Lemma (2.1), we obtain

$$\Big| \sum_k^{(1)} g(k/n) T_{nk}(x) \Big| < K[nx(1-x)]^{-\frac{1}{2}} \sum_k^{(1)} | g(k/n) |.$$

By condition (iii) it is clear that the right-hand side approaches 0 uniformly in $[\delta, 1-\delta]$ as $n \to \infty$. On the other hand, for x in $[0, \delta]$ or $[1-\delta, 1]$ we have

$$\Big| \sum_k^{(1)} g(k/n) T_{nk}(x) \Big| \leqq M \sum_k^{(1)} T_{nk}(x),$$

where M is abound on the absolute value of $g(r)$. For the values of k involved in $\sum_k^{(1)}$ and for the values of x under consideration it is not difficult to see that $| k/n - x | > \delta$. Hence for all n such that $n^{-1/4} < \delta$ we have $\sum_k^{(1)} T_{nk}(x) \leqq \sum_k^* T_{nk}(x)$, where $\sum_k^*$ is taken over those values of k for which $| k - nx | > n^{3/4}$. Consequently, by applying Lemma (2.3), we obtain

$$\Big| \sum_k^{(1)} g(k/n) T_{nk}(x) \Big| < M/(4n^{\frac{1}{2}}),$$

which is valid for all values of n sufficiently large and for all values of x in $[0, \delta]$ and $[1-\delta, 1]$. The proof is therefore complete.

(5.4) Corollary. *If the skeleton $g(r)$ is equivalent to zero and if $g(0) = g(1) = 0$ then the Bernstein polynomials of g approach zero uniformly in $[0, 1]$.*

We return now to the following question. Under what conditions do the Bernstein polynomials of a skeleton in $\mathfrak{S}$ converge uniformly? A complete answer to this question is given in our next theorem.

(5.5) Theorem. *If $f(r)$ is a skeleton in the class $\mathfrak{S}$ then the Bernstein polynomials of $f(r)$ converge uniformly in $[a, b]$ where $0 \leqq a < b \leqq 1$, if and only if the limit function $f_N(x)$ is continuous in $[a, b]$.*

Proof. The necessity of the condition is trivial. Assume then that $f_N(x)$ is continuous in $[a, b]$. We first make the following remark concerning the continuity of $f_N(x)$ at the endpoints of this interval. (In this connection

see footnote [5].) If $a > 0$ then our assumption merely states that $f_N(a+) = f_N(a)$. However, by (3.11) we conclude at once that $f_N(a-) = f_N(a)$, so that $f_N(x)$ is continuous at $x = a$ from the left as well as from the right. An analogous statement holds for $x = b$ in case $b < 1$. Thus in any case we are able to apply Lemma (2.6) and obtain the result that $B_n(f_N; x)$ converges uniformly to $f_N(x)$ in $[a, b]$. Hence in view of the identity

$$B_n(f; x) - f_N(x) = [B_n(f; x) - B_n(f_N; x)] + [B_n(f_N; x) - f_N(x)],$$

the proof will be complete if we can show that $B_n(f; x) - B_n(f_N; x)$ converges to 0 uniformly in $[a, b]$. This will be accomplished by showing that the preceding difference converges to 0 uniformly in the whole interval $[0, 1]$.

Let $g(r) \equiv f(r) - f_N(r)$ for rational r in $[0, 1]$. From (1.1) it is clear that $B_n(g; x) = B_n(f; x) - B_n(f_N; x)$. By Lemma (3.12) we have $f_N(r) \sim f(r)$ and hence $g(r) \sim 0$. Finally, by (1.11) we have $g(0) = g(1) = 0$. We therefore conclude from Corollary (5.4) that $B_n(g; x)$ converges to 0 uniformly in $[0, 1]$. This completes the proof.

We come now to the final question of this section. Under what conditions will the Bernstein polynomials corresponding to two skeletons in $\mathfrak{S}$ converge to the same limit function? This question is answered in the following theorem in which, on account of (1.4), we restrict ourselves to the open interval $(0, 1)$.

(5.6) THEOREM. *If $f(r)$ and $g(r)$ are two skeletons in $\mathfrak{S}$ then $\lim_n B_n(f; x) = \lim_n B_n(g; x)$ for $0 < x < 1$ if and only if $f(r) \sim g(r)$.*

Proof. The proof is immediate. By Theorem (5.1) the limit functions in question are $f_N(x)$ and $g_N(x)$. By Lemma (3.13) $f_N(x) = g_N(x)$ for $0 < x < 1$ if and only if $f(r) \sim g(r)$.

6. Further convergence theorems. It is natural to ask at this point if any interesting convergence theorems for $\{B_n(f; x)\}$ can be obtained by relaxing the restriction that f belong to $\mathfrak{S}$. The Examples (7.1) and (7.3) in **7** below will make it plain that in general no such theorems are to be expected if we stray "too far" from $\mathfrak{S}$. On the other hand, if we stay "close enough" to $\mathfrak{S}$, in the sense of (6.2), (6.6) and (6.9) below, some further results concerning the convergence of $\{B_n(f; x)\}$ can be deduced.

We shall begin with the following result which establishes a necessary condition for convergence.

(6.1) THEOREM. *In order that the Bernstein polynomials of a given bounded skeleton $f(r)$ converge everywhere in $(0, 1)$, or even almost every-*

where, to a function $g(x)$, which is at most simply discontinuous in $[0,1]$, *it is necessary that*

$$(6.2)\qquad \sum_{k=0}^{n}[f(k/n)-g(k/n)]=o(n)\qquad (n\to\infty).$$

Proof. Since the points of discontinuity of $g(x)$ form at most a denumerable set it is clear from the first part of Lemma (2.6) that $B_n(g;x)$ converges to $g(x)$ almost everywhere in $(0,1)$. Hence $B_n(f-g;x)$ approaches 0 almost everywhere in $(0,1)$. Since by (1.3) the sequence $\{B_n(f-g;x)\}$ is uniformly bounded in $[0,1]$, we conclude from the Lebesgue convergence theorem that $\int_0^1 B_n(f-g;x)\,dx=o(1)$. The condition (6.2) now follows from the well known formula

$$(6.3)\qquad \int_0^1 T_{nk}(x)\,dx=1/(n+1)\quad (k=0,1,\cdots,n;\ n=1,2,3,\cdots).$$

(6.4) COROLLARY. *In order that the Bernstein polynomials of a given bounded skeleton $f(r)$ converge to zero almost everywhere in $(0,1)$ it is necessary that $\sum_k f(k/n)=o(n)$ as $n\to\infty$.*

If we strengthen (6.2) by taking the absolute values of the differences $f(k/n)-g(k/n)$ we obtain the sufficient condition (6.6) for the convergence almost everywhere of a subsequence of $\{B_n(f;x)\}$. The comparative weakness of this conclusion is perhaps to be expected since if $f(r)$ is bounded then the left-hand side of (6.6) is certainly $O(n)$.

(6.5) THEOREM. *If for a given skeleton $f(r)$ there exists a skeleton $g(r)$ in $\mathfrak{S}$ such that*

$$(6.6)\qquad \sum_{k=0}^{n}|f(k/n)-g(k/n)|=o(n)\qquad (n\to\infty),$$

then there is a subsequence of $\{B_n(f;x)\}$ which converges to $g_N(x)$ almost everywhere in $(0,1)$.

Proof. From Theorem (5.1) it follows that the sequence $\{B_n(g;x)\}$ converges everywhere in $(0,1)$ to $g_N(x)$. Hence it will suffice to show the existence of a sequence of indices $\{n_i\}$ such that $\lim_i B_{n_i}(f-g;x)=0$ almost everywhere. To do this we start with the simple inequality

$$(6.7)\qquad |B_n(f-g;x)|\leqq\sum_{k=0}^{n}|f(k/n)-g(k/n)|\,T_{nk}(x).$$

This relation and the formula (6.3) immediately yield the inequality

$$\int_0^1 | B_n(f-g;x) | \, dx \leqq [1/(n+1)] \sum_{k=0}^{n} | f(k/n) - g(k/n) | .$$

From the hypothesis (6.6) it follows that the integral on the left is $o(1)$ as $n \to \infty$. Hence by a familiar result [8] there exists a sequence of indices $\{n_i\}$ such that $B_{n_i}(f-g;x)$ approaches 0 almost everywhere in $(0,1)$ as $i \to \infty$. This completes the proof.

Finally, if we strengthen (6.6) to $o(n^{\frac{1}{2}})$ we can prove the convergence of $\{B_n(f;x)\}$ for $0 < x < 1$.

(6.8) THEOREM. *If for a given skeleton $f(r)$ there exists a skeleton $g(r)$ in $\mathfrak{S}$ such that*

$$\sum_{k=0}^{n} | f(k/n) - g(k/n) | = o(n^{\frac{1}{2}}) \qquad (n \to \infty), \tag{6.9}$$

then the sequence $\{B_n(f;x)\}$ converges to $g_N(x)$ for $0 < x < 1$.

Proof. Following the method of the preceding proof we have merely to show that the sequence $\{B_n(f-g;x)\}$ converges to 0 for $0 < x < 1$. But this follows at once from (6.9) by applying Lemma (2.1) to the $T_{nk}(x)$ in (6.7).

(6.10) COROLLARY. *If for a given skeleton $f(r)$ we have $\sum_k | f(k/n) | = o(n^{\frac{1}{2}})$, then $\lim_n B_n(f;x) = 0$ for $0 < x < 1$.*

There is a considerable "gap" between the necessary condition (6.2) and the sufficient condition (6.9). However, by means of Examples (7.5) and (7.8), given in the next section, we establish the fact that this gap cannot be narrowed. Consequently, it is impossible to give a necessary *and* sufficient condition for convergence in terms of the sums involved. Furthermore, by means of examples, which we refrain from including, we have been able to show that in condition (6.2) of Theorem (6.1) the differences on the left side cannot be replaced by their absolute values, while in condition (6.9) of Theorem (6.8) the absolute values of the differences on the left side cannot be replaced by the differences themselves.

7. Examples. This section is devoted to a number of examples which illustrate certain aspects of the theory developed in **4-6**. In this connection we call attention to the remarks made at the end of **4** and at the end of **6**.

---

[8] E. W. Hobson, *The Theory of Functions of a Real Variable*, vol. 2, Cambridge (1926), p. 245.

Throughout this section we shall use the generic notation $r = h/m$, where h and m are relatively prime non-negative integers, to represent rational numbers r in $[0, 1]$ in their lowest terms. Without further emphasis we shall frequently employ the simple fact that $f(r) \equiv f(h/m)$ will appear among the $f(k/n)$ for $k = 0, 1, \cdots, n$ if and only if m is a divisor of n.

(7.1) *Example.* The four quantities occurring in (4.4) and (4.5) may be related in the following manner for $0 < x < 1$,

$$(7.2) \quad \tfrac{1}{2}[\underline{f}_L(x) + \underline{f}_R(x)] = \varliminf_{n\to\infty} B_n(f; x) < \varlimsup_{n\to\infty} B_n(f; x) = \tfrac{1}{2}[\overline{f}_L(x) + \overline{f}_R(x)].$$

Let x_0 be a fixed number in $(0, 1)$ and let a' and a'' be two arbitrary positive numbers. If $r = h/m$ with m odd we define $f(r)$ as a' for r in $(0, x_0)$ and as a'' for r in $(x_0, 1)$. Otherwise we define $f(r)$ as 0. It is seen at once that $\underline{f}_L(x) = \underline{f}_R(x) = 0$ for $0 < x < 1$. Moreover, it is clear that (i) $\overline{f}_L(x) = \overline{f}_R(x) = a'$ for $0 < x < x_0$, (ii) $\overline{f}_L(x) = \overline{f}_R(x) = a''$ for $x_0 < x < 1$, and (iii) $\overline{f}_L(x_0) = a'$ and $\overline{f}_R(x_0) = a''$. Now if n is any power of 2 then $B_n(f; x) = 0$ for $0 < x < 1$. On the other hand, we conclude from Lemmas (2.3) and (2.4) that as $n \to \infty$ through odd values, $B_n(f; x)$ approaches (i) a' if $0 < x < x_0$, (ii) a'' if $x_0 < x < 1$, and (iii) $\tfrac{1}{2}(a' + a'')$ if $x = x_0$. Hence the relation (7.2) follows from Theorem (4.3).

The preceding example illustrates a fact which is not surprising, namely, that the Bernstein polynomials for a badly behaved skeleton may themselves behave very badly. The following example, however, shows that the Bernstein polynomials of a badly behaved skeleton may behave quite well.

(7.3) *Example.* The four quantities occurring in (4.4) and (4.5) may be related in the following manner for $0 < x < 1$,

$$\tfrac{1}{2}[\underline{f}_L(x) + \underline{f}_R(x)] < \varliminf_{n\to\infty} B_n(f; x) = \varlimsup_{n\to\infty} B_n(f; x) < \tfrac{1}{2}[\overline{f}_L(x) + \overline{f}_R(x)],$$

and the convergence implied may even be uniform in $[0, 1]$.

This example depends on the construction of a sequence of rational numbers r_i of the form $h_i/2^i$ with h_i odd $(i = 1, 2, 3, \cdots)$, such that each of the subsequences $\{r_{2j}\}$ and $\{r_{2j-1}\}$ is everywhere dense in $(0, 1)$. We leave it to the reader to convince himself that such a construction is possible. We now define the skeleton $f(r)$ as follows: $f(r_i) = (-1)^i r_i(1 - r_i)$ for all i and $f(r) = 0$ otherwise. It is obvious that $\underline{f}_L(x) = \underline{f}_R(x) = -x(1 - x) < 0$ and $\overline{f}_L(x) = \overline{f}_R(x) = x(1 - x) > 0$ for $0 < x < 1$. We now appeal to Lemma (5.2). Since conditions (i) and (ii) are clearly satisfied it remains

only to consider condition (iii). It will suffice to show that $\sum_{k=0}^{n} | f(k/n) | = O(\log n)$. But this is immediate since the number of non-zero terms in the preceding sum is equal to the exponent of the highest power of 2 that divides n. This exponent is plainly $O(\log n)$. Hence $\lim_n B_n(f;x) = 0$ uniformly in $[0,1]$.

Most of the theory developed in **5** can be illustrated by the following classical example from elementary analysis.

(7.4) *Example.* Let $f(r)$ be defined as follows: $f(0) = f(1) = 0$; when $0 < r < 1$ and $r = h/m$, $f(r) = 1/m$ or, more generally, $f(r) = a_m$ where $\{a_m\}$ is an arbitrary null sequence.

It is at once clear that $f_N(x) = 0$ for $0 \leqq x \leqq 1$ and hence, by Theorems (5.1) and (5.5), $B_n(f;x)$ approaches 0 uniformly in that interval.

The following counter-example will make it clear that in condition (6.2) of Theorem (6.1) the right side cannot be replaced by anything of order smaller than $o(n)$.

(7.5) *Example.* Let c_n be an arbitrary null sequence of non-negative numbers. Then it is possible to find a skeleton $f(r)$ for which (i) $\lim_n B_n(f;x) = 0$ in $(0,1)$, and (ii) $\sum_k f(k/n) \geqq c_n n$ for $n = 2, 3, 4, \cdots$. (See Corollary (6.4).)

In fact we may use for $f(r)$ the skeleton of Example (7.4) with the a_m equal to $2 \operatorname{Max}_{i \geqq m} c_i$. Statement (i) follows immediately from the fact that a_m tends to 0. To verify the truth of (ii) we first observe that if h/m is the reduced form of k/n $(k = 1, 2, \cdots, n-1)$ then $f(k/n) = a_m$. Moreover, since the a_m are monotonically non-increasing we conclude that $f(k/n) \geqq a_n$. Hence we have $\sum_k f(k/n) \geqq (n-1)a_n \geqq 2(n-1)c_n \geqq c_n n$ for $n \geqq 2$.

As our next example we propose to show that in condition (6.9) of Theorem (6.8) the right side cannot be replaced by anything of order larger than $o(n^{\frac{1}{2}})$. Before proceeding with the details of the example itself we first make the following remarks. (a) By evident modifications in the proof [4] of Lemma (2.3) it can be shown that for $0 \leqq x \leqq 1$

$$\Sigma^{\mathrm{I}} T_{nk}(x) > 3/4; \quad \Sigma^{\mathrm{II}} T_{nk}(x) < 1/4, \tag{7.6}$$

where the sums Σ^{I} and Σ^{II} are taken over those values of k for which $| k - nx | \leqq n^{\frac{3}{4}}$ and $| k - nx | > n^{\frac{3}{4}}$, respectively. (b) For any finite set $a_1, a_2, \cdots, a_s$ of real numbers, with each $a_i \geqq 2$, the following inequality can be easily established by induction.

$$a_1 + a_2 + \cdots + a_s \leqq a_1 a_2 \cdots a_s. \tag{7.7}$$

(7.8) *Example.* Let x_0 be any point in $(0,1)$. Then it is possible to find a skeleton $f(r)$ for which (i) $B_n(f;x_0)$ does not approach 0, and (ii) $\sum_k |f(k/n)| = O(n^{\frac{1}{2}})$. (See Corollary (6.10).)

It is clear first of all that there exists a number $P \geqq 5$ such that the interval $I_q \equiv [x_0 - q^{-\frac{1}{2}}, x_0 + q^{-\frac{1}{2}}]$ lies entirely in $(0,1)$ for all $q \geqq P$. For $r = h/m$ we now define $f(r)$ as 1 if m is a prime $p \geqq P$ and $r \in I_p$. We define $f(r)$ as 0 otherwise. From this definition, from the meaning of $\sum^{\mathrm{I}}$, and from (7.6) it follows that $B_p(f;x_0) = \sum^{\mathrm{I}} T_{nk}(x_0) > 3/4$ for all primes $p \geqq P$. This establishes (i). In order to show that (ii) holds, we make the preliminary observation that for any prime $p \geqq P$ the sum $\sum_k f(k/p)$ contains as many non-vanishing terms as there are integers in the interval $[px_0 - p^{\frac{1}{2}}, px_0 + p^{\frac{1}{2}}]$. Since the number of such integers is less than $2p^{\frac{1}{2}} + 1$, we have $\sum_k f(k/p) < 2p^{\frac{1}{2}} + 1 < 3p^{\frac{1}{2}}$. Now let n be an arbitrary positive integer. If n has no prime factors $\geqq P$ then $\sum_k f(k/n) = 0$. In the contrary case let $p_1, p_2, \cdots, p_s$ be the distinct primes $\geqq P$ which divide n. We may then write the sum $\sum_k f(k/n)$ in the form $\sum_{i=1}^{s} \sum_j f(j/p_i)$ which, by the remark made above, is less than $3\sum_i p_i^{\frac{1}{2}}$. But the latter by (7.7) does not exceed $3\prod_i p_i^{\frac{1}{2}}$ since $p_i^{\frac{1}{2}} \geqq P^{\frac{1}{2}} > 2$. We conclude therefore that $\sum_k f(k/n) < 3n^{\frac{1}{2}}$ for all values of n, which implies (ii).

The preceding example is, of course, open to the criticism that it establishes the non-convergence of $\{B_n(f;x)\}$ only at a single point x_0. It is possible, however, to modify the definition of $f(r)$ in such a way that $\{B_n(f;x)\}$ fails to approach 0 for any value of x in $(0,1)$, while (ii) remains satisfied. Since the details of this modification are moderately complicated we have contented ourselves with the simpler example above.

(7.9) *Example.* We define a skeleton $f(r)$ as follows. Let $f(r) = 1$ for r in (a,b) where $0 < a < b < 1$, and let $f(r) = 0$ otherwise. It is evident that f belongs to $\mathfrak{S}$ and that for its normalized extension we have (a) $f_N(x) = 0$ for $0 \leqq x < a$ and for $b < x \leqq 1$, (b) $f_N(x) = 1$ for $a < x < b$, and (c) $f_N(a) = f_N(b) = \frac{1}{2}$. By Theorem (5.1) the sequence $\{B_n(f;x)\}$ converges to $f_N(x)$ for all x in $[0,1]$. Furthermore, by Theorem (5.5) this convergence is uniform in any closed subinterval of $[0,1]$ which contains neither a nor b.

Let us now assign to x a fixed value p and set $q = 1 - p$. Then $B_n(f;p)$ assumes the form

$$\sum_{na<k<nb} C_{n,k} p^k q^{n-k}. \tag{7.10}$$

In view of the preceding discussion we see that the sum (7.10) approaches the limit (i) 0 if $0 \leqq p < a$ or $b < p \leqq 1$, (ii) 1 if $a < p < b$, (iii) $\frac{1}{2}$ if $p = a$ or $p = b$. If we interpret p as the constant probability of an event E then the sum (7.10) is precisely the probability that the relative frequency of the occurrence of E in n trials lies between a and b. Consequently, the results (i)-(iii) for the value of the limit of the sum (7.10) have evident interpretations in probability. In particular, if (a, b) is regarded as an arbitrary neighborhood of p then the result (ii) is exactly the weak law of large numbers. (As a matter of fact this law can also be deduced from the first part of Lemma (2.6) the proof of which does not involve the central limit theorem.) It is therefore conceivable that the theory developed in this paper may find other applications to questions in probability.

MICHIGAN STATE COLLEGE.

ASYMPTOTIC EQUILIBRIA.*

By AUREL WINTNER.

1. A class of standard problems concerning the asymptotic behavior of solutions of linear differential equations centers around a simple "Abelian" theorem, which may be formulated as follows (cf. [3]): If (f_{ik}), where $i, k = 1, \cdots, n$, is a matrix of continuous functions $f_{ik} = f_{ik}(t)$, $0 \leqq t < \infty$, satisfying the n^2 conditions

$$\int^{\infty} | f_{ik}(t) | \, dt < \infty, \tag{1}$$

then each of the n components $x_i(t)$ of every solution of the linear differential equations

$$x_i' = \sum_{k=1}^{n} f_{ik}(t) x_k \tag{2}$$

tends to a finite limit as $t \rightarrow \infty$. In view of the primitive nature of this Abelian fact, it can be expected that some corresponding result holds in the general case of non-linear differential equations

$$x_i' = f_i(t; x_1, \cdots, x_n), \tag{3}$$

where $f_1, \cdots, f_n$ are continuous functions of position on the $(n+1)$-dimensional region

$$0 \leqq t < \infty; \qquad -\infty < x_1 < \infty, \cdots, -\infty < x_n < \infty \tag{4}$$

and satisfy an additional restriction which, corresponding to (1), compels the existence of n finite limits $x_i(\infty)$ for every solution $x_i(t)$ of (3). In other words, *every* solution *path* $x_i = x_i(t)$ should tend, as $t \rightarrow \infty$, to a certain *point* of the x-space, a point of asymptotic equilibrium.

The relevant Abelian theorem delimiting this situation will turn out to be as follows:

(i) *Let n real-valued, continuous functions f_i of the position $(t; x_1, \cdots, x_n)$ on the $(n+1)$-dimensional region* (4) *be such that*

$$\int^{\infty} | f_i(t; 0, \cdots, 0) | \, dt < \infty \tag{5}$$

* Received May 26, 1945.

and that there exists a matrix (λ_{ik}) *of* n^2 *non-negative, measurable functions*

$$(6) \qquad \lambda_{ik} = \lambda_{ik}(t), \qquad 0 \leqq t < \infty$$

satisfying

$$(7) \qquad \int^{\infty} \lambda_{ik}(t)\,dt < \infty$$

and

$$(8) \quad | f_i(t; x_1^*, \cdots, x_n^*) - f_i(t; x_1^{**}, \cdots, x_n^{**}) | \leqq \sum_{k=1}^{n} \lambda_{ik}(t) | x_k^* - x_k^{**} |,$$

where t, x_k^*, x_k^{**} *are arbitrary. Suppose, in addition, that the functions* $\lambda_{ik}(t)$ *are bounded for* $0 \leqq t < T$ *if* $T < \infty$ *(but not necessarily if* $T = \infty$*). Then each of the* n *components* $x_i(t)$ *of a solution of the differential equations* (3) *tends to a finite limit,* $x_i(\infty)$, *as* $t \to \infty$, *no matter what the* n *initial values* $x_i(0)$ *determining the solution may be.*

Thus, $(x_1(0), \cdots, x_n(0)) \to (x_1(\infty), \cdots, x_n(\infty))$ is a mapping defined on the whole x-space. It is a question of stability whether the inverse of this mapping exists on the *whole* x-space and is single-valued and continuous. In this connection, cf. [1].

It may be mentioned that, in virtue of (8) and (7), condition (5) is equivalent to

$$(9) \qquad \int^{\infty} | f_i(t; x_1^0, \cdots, x_n^0) | \,dt < \infty,$$

where $(x_1^0, \cdots, x_n^0)$ is an arbitrary point of the x-space. In order to see this, it is sufficient to choose $x_k^* = x_k^0$, $x_k^{**} = 0$.

2. The integral (5) is 0 if

$$(10) \qquad f_i(t; x_1, \cdots, x_k) = \sum_{k=1}^{n} f_{ik}(t) x_k,$$

that is, if (3) is the linear system (2). It is also clear that condition (8) then is satisfied by the continuous functions $\lambda_{ik}(t) = | f_{ik}(t) |$; so that (7) becomes precisely (1). Hence, the fact quoted after (1) is contained in (i).

Correspondingly, an application made in [3] of the linear case, (2), of (i) can now be extended to non-linear differential equations, as follows:

(ii) *Let a matrix* (c_{ik}) *of* n^2 *real constants* $c_{11}, \cdots, c_{nn}$ *be such as to have linear elementary divisors and purely imaginary characteristic numbers only (so that the general solution of the differential equations*

$$(11) \qquad y_i' = \sum_{k=1}^{n} c_{ik} y_k$$

is represented by n almost-periodic polynomials

$$y_i(t) = \sum_{k=1}^{n} \alpha_{ik} \cos (\mu_k t + \beta_{ik}), \tag{12}$$

where $-|\mu_1|^2, \cdots, -|\mu_n|^2$ *are the squares of the characteristic numbers of* (c_{ik}) *and* α_{ik}, β_{ik} *denote* $2n^2$ *integration constants, only n of which are independent). Then, if* $f_1, \cdots, f_n$ *are functions satisfying the assumptions of* (i), *there belongs to every solution* $(x_1(t), \cdots, x_n(t))$ *of the differential equations*

$$x_i' = \sum_{k=1}^{n} c_{ik} x_k + f_i(t; x_1, \cdots, x_n) \tag{13}$$

a solution (12) *of the trivial approximation,* (11), *to* (13) *in such a way that*

$$x_i(t) - y_i(t) \to 0 \text{ as } t \to \infty, \tag{14}$$

where $i = 1, \cdots, n$.

Since 0 is a purely imaginary number, the algebraic assumptions of (ii) are satisfied if (c_{ik}) is the zero matrix. In this case, every component $y_i(t)$ of every solution (12) of (11) is constant. Hence, the assertion, (14), of (ii) becomes the assertion of (i). Since (13) becomes (3), it follows that (ii) is a generalization of (i).

Actually, the more elaborate theorem, (ii), is equivalent to the purely Abelian lemma, (i). In fact, (ii) can be reduced to (i) by the same device of "variation of constants" which, in [3], reduced the linear case of (ii) to the linear case of (i). Inasmuch as not even the details of this reduction need a rewording, it will be sufficient to prove (i).

3. Even if the functions f_i are of class C' (or, for that matter, regular-analytic) on the whole region (4), a solution $x_i(t)$ of (3) which is determined by a set of n initial values $x_i(0)$ will not, in general. exist except for small t. In fact, the general existence theorem of non-linear differential equations is purely local in nature (even in the regular-analytic case). Hence, *the unrestricted existence of every solution* $x_i(t)$ *of* (3), (13) *throughout the whole range* $0 \leqq t < \infty$ *is part of the statements of* (i), (ii).

Correspondingly, the proof of (i) will fall into two parts. That part dealing with the unrestricted existence of the solutions will be taken care of by (iv) below. The remaining part, that concerning the existence of the limits $x_i(\infty)$, can then be formulated in a manner sharper than what is claimed by (i); namely, as follows:

(iii) *Let* $f_1, \cdots, f_n$ *satisfy the assumptions of* (i) *except its additional*

assumption made after (8); *so that the functions* $\lambda_{ik}(t)$ *need not now be bounded on bounded t-intervals. The remaining assumptions of* (i) *imply the existence of a fixed value* T *having the following property: Any solution* $(x_1(t), \cdots, x_n(t))$ *of* (3) *which exists at* $t = T$ (*for instance, any solution of* (3) *satisfying the initial conditions*

$$(15) \qquad x_i(T) = x_i^0,$$

where $x_1^0, \cdots, x_n^0$ *are arbitrary*) *can be prolonged for every* t *past* T, *and all* n *components* $x_i(t)$, $T \leqq t < \infty$, *of the resulting solutions of* (3) *will tend to finite limits* $x_i(\infty)$.

That assumption of (i) which is omitted in (iii) means the existence of a Lipschitz constant $L = L_T$ satisfying

$$(16) \qquad | f_i(t; x_1^*, \cdots, x_n^*) - f_i(t; x_1^{**}, \cdots, x_n^{**}) | \leqq L \sum_{k=1}^{n} | x_k^* - x_k^{**} |$$

for unspecified x_k^*, x_k^{**} if $0 \leqq t \leqq T$, where T is arbitrarily fixed ($< \infty$). Hence, in order to reduce (i) to (iii), it is sufficient to choose T so large as required by (iii) and then to apply, on the resulting ranges $T \leqq t < \infty$ and $0 \leqq t \leqq T$, the respective assertions of (iii) and of the following lemma:

(iv) *Let* n *real-valued, continuous functions of position on an* $(n+1)$-*dimensional region*

$$(17) \qquad 0 \leqq t \leqq T; \qquad -\infty < x_1 < \infty, \cdots, -\infty < x_n < \infty$$

be such as to satisfy a uniform Lipschitz condition (16), *where* $0 \leqq t \leqq T$ (*and* x_k^*, x_k^{**} *are arbitrary*). *Then the solution* $x_i(t)$ *of* (3) *determined by any set of initial values* $x_i(0)$ *exists not only for "sufficiently small"* t *but on the whole t-range,* $0 \leqq t \leqq T$, *admitted in* (17).

This lemma, which is undoubtedly well-known, is of independent interest, since it assures existence in the large for all solutions. It may be verified, for instance, as follows: An application of (16) to $x_k^* = x_k$, $x_k^{**} = 0$ gives

$$| f_i(t; x_1, \cdots, x_n) | \leqq L \sum_{k=1}^{n} | x_k | + | f_i(t; 0, \cdots, 0) |,$$

where L is constant on the region (17). Since $f_i(t; 0, \cdots, 0)$ is a continuous, hence bounded, function on the interval $0 \leqq t \leqq T$, it follows that there exists a constant satisfying

$$| f_i(t; x_1, \cdots, x_n) | \leqq \text{const.} \max (| x_1 |, \cdots, | x_n |, 1)$$

on the region (17). Hence, (iv) is contained in the criteria proved in [2].

4. Since (iii) and (iv) imply (i), and since (ii) is equivalent to (i), only (iii) remains to be proved.

In the assumptions of (iii), the functions $\lambda_{ik}(t)$ occurring in (8) can be unbounded on every interval $t_0 < t < t_0 + \epsilon$ (in fact, they can be ∞ on a dense t-set, and remain ∞ on a dense t-set after *any* alteration of their values on t-sets of measure 0), since only the L-integrability of the functions $\lambda_{ik}(t)$ is required. In particular, (iii) does not assume any local Lipschitz condition (16) (for bounded t-ranges). Nevertheless, (iii) can be proved by the method of successive approximations. This will be shown by an adaptation of that "accruing" variant of the usual estimate of successive approximations on which Lichtenstein's fundamental existence theorems for the movement of gravitating fluid masses are based (conditions (8), (7) there correspond to such data as a Hölder condition and the behavior of potential or velocity fields at infinity).

According to (7),

$$\sum_{k=1}^{n} \int_{T}^{\infty} \lambda_{ik}(t)\,dt \leqq \theta < 1 \qquad (i = 1, \cdots, n), \tag{18}$$

if $T = T_\theta$ is large enough. It will be shown that every T satisfying (18) for some $\theta = \theta_T < 1$ is a T having the property claimed by (iii).

For fixed θ and T, let t be restricted to the interval $T \leqq t < \infty$. Then all solutions of (3) can be thought of as belonging to initial conditions (15), where the n constants x_i^0 are arbitrary. Let these initial values be fixed. According to (15), the corresponding successive approximations are defined by the recursion formula

$$x_i^{m+1}(t) = x_i^0 + \int_{T}^{t} f_i(u; x_1^m(u), \cdots, x_n^m(u))\,du, \tag{19}$$

where $m = 0, 1, \cdots$ and

$$x_i^0(t) = x_i^0 \text{ for all } t \tag{20}$$

$(i = 1, \cdots, n)$. This defines continuous functions $x_1^m(t), \cdots, x_n^m(t)$ for every t and for every m, since the functions $f_i(t; x_1, \cdots, x_n)$ are defined and continuous on the whole range (4).

Let C_m denote the least common upper bound of the n functions

$$| x_1^{m+1}(t) - x_1^m(t) |, \cdots, | x_n^{m+1}(t) - x_n^m(t) | \tag{21}$$

on the interval $T \leqq t < \infty$ (so that, for the present, $C_m \leqq \infty$). Since (20) and the case $m = 0$ of (19) imply that

$$| x_i^1(t) - x_i^0(t) | \leqq \int_T^t | f_i(u; x_1^0, \cdots, x^0) | \, du,$$

the formulation (9) of the assumption (5) of (iii) assures that $C_1 < \infty$. On the other hand, if m in (19) is replaced by $m-1$, it follows by subtraction that

$$| x_i^{m+1}(t) - x_i^m(t) | \leqq \int_T^t \sum_{k=1}^n \lambda_{ik}(u) \, | x_k^m(u) - x_k^{m-1}(u) | \, du,$$

by the case $x_k^* = x_k^m(u)$, $x_k^{**} = x_k^{m-1}(u)$ of (8). But the definition of C_m shows that the last integral is majorized by

$$\int_T^t \sum_{k=1}^n \lambda_{ik}(u) C_{m-1} du \leqq C_{m-1} \sum_{k=1}^n \int_T^\infty \lambda_{ik}(u) \, du,$$

a value which, according to (18), does not exceed $C_{m-1}\theta$. Hence, if the definition of C_m is compared with the last two formula lines, it follows that $C_m \leqq C_{m-1}\theta$. In view of $C_1 < \infty$, this means that there exists a constant $C_0 = C_0(\theta)$ satisfying $C_m \leqq C_0 \theta^m$ for every m.

Since the constant C_m majorizes the n functions (21) for $T \leqq t < \infty$, the inequality $C_m \leqq C_0\theta^m$, where $\theta < 1$ by assumption, implies that the n series

$$\sum_{m=0}^\infty \{ x_i^{m+1}(t) - x_i^m(t) \}$$

are uniformly convergent on the interval $T \leqq t < \infty$ and represent there n bounded functions. It follows therefore from the constancy of the functions (20) that, as $m \to \infty$, each of the n functions $x_i^m(t)$ tends to a limit function uniformly for $T \leqq t < \infty$ and that, if $x_i(t)$ denotes the limit function, $x_i(t)$ remains bounded as $t \to \infty$.

In particular, uniform convergence takes place on every bounded portion of the range $T \leqq t < \infty$. This in itself implies that $x_i(t)$ represents a solution of (3) for $T \leqq t < \infty$. Hence, the proof of (iii) will be complete if it is shown that the n limits $x_i(\infty)$ exist for the solution just constructed.

To this end, choose $x_k^* = x_k(t)$ and $x_k^{**} = x_k^0$ in (8). This gives

$$| f_i(t; x_1(t), \cdots, x_n(t)) | \leqq | f_i(t; x_1^0, \cdots, x_n^0) | + \sum_{k=1}^n \lambda_{ik}(t) \, | x_k(t) - x_k^0 |.$$

But the functions $x_1(t), \cdots, x_n(t)$ were seen to be bounded for $T \leqq t < \infty$. Hence, the function multiplying $\lambda_{ik}(t)$ on the right of the last formula line

remains bounded as $t \to \infty$, every x_k^0 being a constant. Consequently, from the last formula line and from (9) and (7),

$$\int^{\infty} | f_i(t; x_1(t), \cdots, x_n(t)) | \, dt < \infty.$$

Since $(x_1(t), \cdots, x_n(t))$ is a solution of (3), this means that the n integrals $\int^{\infty} x_i'(t) dt$ are absolutely convergent. In particular, they are convergent. But their convergence is equivalent to the existence of finite limits $x_i(\infty)$, since $x_i' = dx_i/dt$.

5. The above formulations and proofs are within the real field. However, it is clear from the proofs that *the results can immediately be transcribed to the analytic case of the complex field*, where the real t-axis is replaced by the complex z-plane. What then corresponds to the half-line $T < t < \infty$ may, for instance, be a wedge $| \arg z | < \alpha$, and what corresponds to $t \to \infty$ is the limit process $z \to \infty$, where z is restricted to the wedge.

Needless to say, the theorems which thus result for a (partial) neighborhood of $z = \infty$ can be transformed into corresponding theorems relating to the behavior of arbitrary solutions of certain non-linear analytic differential equations near a singular point at $z = 0$.

Addendum (May 30, 1945). In correcting an impression I originally had concerning the possible instability of the mapping mentioned after (i), Professor Siegel observed that, *under the assumptions of* (i), *the mapping* $(x_1(0), \cdots, x_n(0)) \to (x_1(\infty), \cdots, x_n(\infty))$ *of the x-space is continuous and one-to-one.* The proof is published here with his kind permission.

Let $u \cdot v = u_1 v_1 + \cdots + u_n v_n$ and $| u | = (u \cdot u)^{\frac{1}{2}}$, where $u = x, f, \cdots$. Then (3) and (8) become $x' = f(t; x)$ and

$$| f(t; x^*) - f(t; x^{**}) | \leqq \lambda(t) | x^* - x^{**} |, \text{ where } \int^{\infty} \lambda(t) dt < \infty,$$

by (7). Let

$$r = r(t) = | x^*(t) - x^{**}(t) |,$$

where $x = x^*(t)$ and $x = x^{**}(t)$ are two solutions of $x' = f(t; x)$. Then, by Schwarz's inequality,

$$| rr' | \leqq r \, | \{x^*(t) - x^{**}(t)\}' | = r \, | f(t; x^*(t)) - f(t; x^{**}(t)) |.$$

Since the last difference is majorized by $\lambda(t)r$, it follows that $|r'|$ cannot exceed $r\lambda$ for any t. It follows, therefore, from the absolute integrability of $\lambda(t)$ that the limit, $r(\infty)$, of $r(t)$ as $t \to \infty$ has a value contained between two bounds $r(0)e^{\pm c}$, where c is independent of $r(0)$. In other words,

$$|x^*(0) - x^{**}(0)|\, e^{-c} \leqq |x^*(\infty) - x^{**}(\infty)| \leqq |x^*(0) - x^{**}(0)|\, e^{c},$$

where the existence of the limits is assured by (i).

According to (i), the mapping $x(0) \to x(\infty)$ is defined on the whole x-space. Let this mapping be denoted by μ. Then the last formula line implies that μ is continuous and has a unique inverse μ^{-1}. Hence, all that remains to be shown is that M is the whole x-space, if M denotes the μ-image of the whole x-space.

First, M is an open set, by Jordan's theorem. Next, if $x^1(\infty), x^2(\infty), \cdots$ is a convergent sequence of points of M, the last formula line implies that $x^1(0), x^2(0), \cdots$ is a bounded sequence and that, if $x(0)$ denotes a cluster point of the latter sequence, then $x(0)$ is identical with the point $\lim x^n(\infty)$ the existence of which is assumed. Accordingly, the set M is closed. Since it is open, it must be the whole space.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "Integrability in the large and dynamical stability," *American Journal of Mathematics*, vol. 65 (1943), pp. 273-278.

[2] A. Wintner, "The non-local existence problem of ordinary differential equations," *ibid.*, vol. 67 (1945), pp. 277-284.

[3] ———, "Small perturbations," *ibid.*, pp. 417-430.

MONOTONE SERIES.*

By R. W. Hamming.

1. Introduction. Let Σc_n be a monotone convergent series and let Σd_n be a monotone divergent series. Pringsheim has shown, contrary to first impressions, that there exist monotone series such that

$$\limsup c_n/d_n = \infty.$$

On the other hand

$$\liminf c_n/d_n = 0,$$

since otherwise $c_n/d_n \geq \epsilon$ for all n would imply $\Sigma c_n \geq \epsilon \Sigma d_n$.

The object of the present paper is to study more closely the possible behavior of c_n/d_n as n approaches infinity, in particular, the question of how many of the terms of c_n/d_n may be bounded away from zero. By replacing Σc_n by $1/\epsilon \Sigma c_n$ this is transformed into the question of how many of the c_n may be greater than the corresponding d_n. Since this may happen infinitely often, we study the function $A(n)$ which is defined as the average number of times $c_n \geq d_n$.

The basic result is given in Theorem 1 of **2**. Section **3** contains an elementary example which shows that not only may c_n be greater than or equal to d_n infinitely often, but that for certain series

$$\limsup A(n) = 1.$$

Section **4** discusses the summability of the function $A(n)$.

2. Fundamental theorem. Throughout this paper Σc_n is a monotone convergent series and Σd_n is a monotone divergent series. Let $N(n)$ be the number of terms for which $c_i \geq d_i$, $i = 1, 2, \cdots, n$. Then the function

$$A(n) = N(n)/n$$

measures the average number of times the c_i are at least as large as the corresponding d_i terms.

Theorem 1. *For every $\epsilon > 0$ and any $k > 1$ there exist infinitely many r_i $(r_i \to \infty)$ such that throughout the interval $r_i \leq n \leq kr_i$ we have $A(n) < \epsilon$.*

To prove the theorem we first develop two lemmas.

* Received April 9, 1945.

LEMMA 1. *If, for a given positive ϵ less than 1, $A(n) \geq \epsilon$ for some value or values of n in each of the intervals*

$$K^m \leq n \leq K^{m+\frac{1}{2}} \quad \textit{and} \quad K^{m+\frac{1}{2}} \leq n \leq K^{m+1}$$

where $K \geq 4/\epsilon^2$, then for some value $n = n_m$ in the combined interval

$$K^m \leq n \leq K^{m+1} \tag{2.1}$$

both

$$A(n_m) \geq \epsilon \quad \textit{and} \quad c_{n_m} \geq d_{n_m}. \tag{2.2}$$

Proof. If the first interval begins with $A(n) < \epsilon$ then the first n for which $A(n) \geq \epsilon$ will provide a suitable n_m. The same argument can be applied to the second interval, so that the lemma is proved except in the case where the first interval begins and ends with $A(n) \geq \epsilon$. In this case there must be a suitable n_m in the first interval since otherwise we would have at the end of the interval

$$A(K^{m+\frac{1}{2}}) = N(K^{m+\frac{1}{2}})/K^{m+\frac{1}{2}} = N(K^m)/K^{m+\frac{1}{2}} \leq K^m/K^{m+\frac{1}{2}} = 1/K^{\frac{1}{2}} \leq \epsilon/2.$$

LEMMA 2. *For every positive ϵ less than 1 there exist infinitely many intervals*

$$K^m \leq n \leq K^{m+\frac{1}{2}} \quad \textit{and} \quad K^{m+\frac{1}{2}} \leq n \leq K^{m+1} \qquad (K \geq 4/\epsilon^2)$$

throughout which $A(n) < \epsilon$.

Proof. We first assume, contrary to what we wish to prove, that in every interval $A(n) \geq \epsilon$ for some value or values, and we apply Lemma 1 to conclude that in the combined interval (2.1) there is an n_m such that both parts of (2.2) hold. Thus, selecting every other suitable n value, that is only n_{2m}'s, and using the monotoneity of the c_n we have

$$\begin{aligned} \sum_n c_n &\geq \sum_m c_{n_{2m}}\{N(n_{2m}) - N(n_{2m-2})\} \\ &\geq \sum_m d_{n_{2m}}\{N(n_{2m}) - N(n_{2m-2})\}. \end{aligned} \tag{2.3}$$

We now introduce the function δ_m defined by

$$\delta_m = \frac{N(n_{2m}) - N(n_{2m-2})}{N(n_{2m+2}) - N(n_{2m})}.$$

By definition $A(n) = N(n)/n \leq 1$, so that $A(n_{2m}) \geq \epsilon$ becomes

$$n_{2m} \geq N(n_{2m}) \geq \epsilon n_{2m}.$$

Applying this to δ_m we have

$$\delta_m \geq \frac{\epsilon n_{2m} - n_{2m-2}}{n_{2m+2}}.$$

But n_{2m} satisfies $K^{2m} \leq n_{2m} \leq K^{2m+1}$ so that

$$\delta_m \geq \frac{\epsilon K^{2m} - K^{2m-1}}{K^{2m+3}} \geq \frac{\epsilon K - 1}{K^4} \geq \frac{4/\epsilon - 1}{K^4} \geq \frac{3}{K^4} > 0.$$

Returning to (2. 3) and using the result for δ_m as well as the monotoneity of the d_n we have

$$\sum_n c_n \geq \sum_m d_{n_{2m}} \delta_m \{N(n_{2m+2}) - N(n_{2m})\} \geq (3/K^4) \sum_{n=n_2} d_n$$

which is a contradiction.

If now we assume that a finite number of the original intervals contained no suitable n values, this would lead to at most a finite number of the double intervals failing to have the properties (2. 2). By deleting a finite number of terms from Σd_n to correspond to these exceptions the divergence is not affected, and hence the contradiction will still occur. Thus Lemma 2 is proved. With a change in notation it is easily seen that Theorem 1 is a special case of Lemma 2.

COROLLARY to Theorem 1. *If $A(n)$ approaches a limit, then this limit is zero.*

3. An example. The corollary discusses the case where $A(n)$ approaches a limit. That $A(n)$ need not approach a limit is shown by the following example. Indeed the example shows that while according to Theorem 1 $A(n)$ must become arbitrarily small throughout arbitrarily long intervals, nevertheless the lim sup of $A(n)$ may be 1.

For the monotone divergent series let $d_n = 1/(n \log n)$. For Σc_n set $c_n = \dfrac{2m}{2^{2^m} \log 2^{2^m}}$ whenever $2^{2^{(m-1)}} < n \leq 2^{2^m}$. Clearly Σc_n is monotone. The proof of its convergence runs as follows:

$$\Sigma c_n = \Sigma (2^{2^m} - 2^{2^{(m-1)}}) \frac{2m}{2^{2^m} \log 2^{2^m}} \leq \frac{2}{\log 2} \Sigma \frac{m}{2^m} .$$

We note that, as in Pringsheim's example, $\limsup c_n/d_n = \infty$ since for $n = 2^{2^m}$ we have $c_n/d_n = 2m$.

For n satisfying $\dfrac{1}{m} \cdot 2^{2^m} \leq n \leq 2^{2^m}$ we have

$$\begin{aligned} \frac{c_n}{d_n} &\geq \frac{2m}{2^{2^m} \log 2^{2^m}} \cdot \frac{2^{2^m}}{m} \log\left(\frac{2^{2^m}}{m}\right) \\ &\geq 2\left\{1 - \frac{\log m}{2^m \log 2}\right\} \geq 1 \qquad (m \geq 1). \end{aligned}$$

Now calculating $A(n)$ at $n = 2^{2^m}$ we have

$$A(2^{2^m}) \geq \frac{(1-(1/m))2^{2^m}}{2^{2^m}} = 1 - \frac{1}{m}$$

and the lim sup $A(n) = 1$.

4. Summability of $A(n)$. From the two given monotone series we have derived a sequence $A(n)$ which, as we have shown in **2** and **3**, may either approach zero or oscillate as n approaches infinity. One of the standard methods of treating an oscillating sequence is to calculate its Cesàro-Hölder means. Since the Cesàro and Hölder methods are equivalent we may use the Hölder method in proving the following general result.

THEOREM 2. *For every $\epsilon > 0$ and any $k > 1$ there exist infinitely many r_i $(r_i \to \infty)$ such that throughout the interval $r_i \leq n \leq kr_i$ we have $C^{(m)}(n) < \epsilon$, where $C^{(m)}(n)$ is the m-th Cesàro-Hölder mean.*

Proof. The proof is by induction with Theorem 1 as a basis, m being zero. Thus we assume that we can pick any E and K and find infinitely many R_i $(R_i \to \infty)$ such that

$$C^{(m-1)}(n) < E \quad \text{for} \quad R_i \leq n \leq KR_i.$$

Pick $E = \epsilon/2$, $K = 2k/\epsilon$, and set $\alpha = 1/k$. Thus, since $C^{(m-1)}(n) \leq 1$ for all n,

$$C^{(m)}(\alpha KR_i) \leq \frac{R_i + E(\alpha K - 1)R_i}{\alpha KR_i} = \frac{1}{\alpha K} + E\left(1 - \frac{1}{\alpha K}\right)$$
$$< \epsilon/2 + \epsilon/2 = \epsilon$$

and we have for all n satisfying $\alpha KR_i \leq n \leq KR_i$

$$C^{(m)}(n) < \epsilon.$$

Setting $R_i = \epsilon r_i/2$ the above interval becomes

$$\alpha KR_i = \frac{1}{k} \cdot \frac{2k}{\epsilon} \cdot \frac{\epsilon r_i}{2} \leq n \leq \frac{2k}{\epsilon} \cdot \frac{\epsilon r_i}{2} = KR_i$$

or

$$r_i \leq n \leq kr_i \qquad (r_i \to \infty).$$

COROLLARY to Theorem 2. *If $A(n)$ is summable $C^{(m)}(n)$ then the sum is zero.*

In closing it is to be noted that examples, only slightly more complicated than that in **3**, can be given to show that $A(n)$ may be summable $C^{(m)}(n)$ and not $C^{(m-1)}(n)$, and thus Theorem 2 is not vacuous.

UNIVERSITY OF ILLINOIS,
URBANA, ILLINOIS.

CONFORMAL MAPS WITH ISOTHERMAL SYSTEMS OF SCALE CURVES.*

By John De Cicco.

1. **Introduction.** Let a surface Σ be mapped in a point-to-point fashion upon a plane π. The *scale function* $\sigma = ds/dS$, which is the ratio of the differentials of arc lengths on the plane π and the surface Σ, depends upon the position of the point and also upon the direction through the point. It is independent of the direction through the point if, and only if, the mapping of Σ upon π is conformal. The scale σ is a constant only in the degenerate situation where Σ is developable and the mapping is an unrolling of Σ upon π, followed by a similitude in π.

A *scale curve* is the locus of a point along which the scale σ does not vary. For a non-conformal map of Σ upon π, there are ∞^2 scale curves. Under a conformal map, there are ∞^1 scale curves. In the degenerate situation mentioned above, every curve is a scale curve. Henceforth we shall exclude this case from consideration so that the scale function σ is a non-constant function.

In the present paper, we shall be concerned chiefly with conformal maps. Hence σ is a non-constant function which depends only upon the position of the point; and there are ∞^1 scale curves. It is remarked that any family of ∞^1 scale curves in the plane π may represent the scale curves of a conformal map of some surface Σ upon the plane π. In particular, any isothermal system can be the scale curves of a conformal map of some surface Σ upon π. We shall consider the surfaces Σ applicable upon a surface of revolution for which there exists a conformal map of Σ upon the plane π with an isothermal family of scale curves, which are neither parallel straight lines nor concentric circles.

Since we shall be mainly concerned with surfaces applicable upon surfaces of revolution, it is convenient to introduce the following terminology. A *sphere-like surface* is any surface of constant, but not zero, gaussian curvature. A *vase-like surface* is any surface which is applicable upon a surface of revolution of variable gaussian curvature. Thus a surface which is applicable upon a surface of revolution may be either developable, or sphere-like, or vase-like. These three classes are mutually exclusive.

Obviously there always exists a developable surface Σ for which any isothermal system of curves may represent the scale curves of a conformal map

* Received March 5, 1945. Presented to the American Mathematical Society, August, 1944.

of Σ upon the plane π. We shall omit this case from consideration so that henceforth the gaussian curvature is assumed to be not identically zero.

Of course, any sphere-like or vase-like surface Σ may be conformally mapped upon the plane π such that the scale curves form an isothermal family. For any such Σ the scale curves are parallel straight lines or concentric circles. For a sphere-like surface we have proved that these are the only possible conformal representations on π with an isothermal family of scale curves. *The main problem of this paper is to discover the forms of the linear-elements of all vase-like surfaces Σ for which there exist conformal maps upon a plane π with isothermal systems of scale curves which are neither parallel straight lines nor concentric circles.* It is found that there are essentially *seven* types in the real domain, or *six* types in the complex domain.

For our work, we shall obtain, in the case of a conformal representation, the analytic form of a theorem of Kasner which states that a surface is applicable upon a surface of revolution if, and only if, it possess an isothermal system of geodesics.[1]

2. Preliminary formulas. Consider any conformal map of a surface Σ upon a plane π with minimal coordinates $(u = x + iy,\ v = x - iy)$. The linear-element of Σ is

$$ds^2 = 2e^{\lambda(u,v)}dudv. \tag{1}$$

The gaussian curvature G is defined by the formula

$$G = - e^{-\lambda}\lambda_{uv}, \tag{2}$$

and its partial derivatives of the first order with respect to u and v are

$$G_u = - e^{-\lambda}(\lambda_{uuv} - \lambda_u\lambda_{uv}), \qquad G_v = - e^{-\lambda}(\lambda_{uvv} - \lambda_v\lambda_{uv}). \tag{3}$$

By (2), it is seen that the surface Σ is developable if, and only if, λ is harmonic, that is, $\lambda_{uv} = 0$. From (3), it follows that Σ is sphere-like if, and only if, λ satisfies the system of two partial differential equations of the third order: $\lambda_{uuv} - \lambda_u\lambda_{uv} = 0$, $\lambda_{uvv} - \lambda_v\lambda_{uv} = 0$.

The differential equation of the ∞^2 geodesics of the surface Σ is

$$v'' = v'(\lambda_u - v'\lambda_v). \tag{4}$$

3. The analytic conditions in a conformal representation for a vase-

[1] Kasner, "Isothermal systems of geodesics," *Transactions of the American Mathematical Society*, vol. 5 (1904), pp. 55-60. See also De Cicco, "New proofs of the theorems of Beltrami and Kasner on linear families," *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 407-412.

like surface. The linear-element (1) represents a vase-like surface Σ if, and only if, λ is given by an equation of the form

$$\lambda = F[\phi(u) + \psi(v)] + \log \phi_u \psi_v, \tag{5}$$

where $\phi_u \neq 0$ and $\psi_v \neq 0$. This means that $[\phi + \psi]$ and $[\lambda - \log \phi_u \psi_v]$ are functionally dependent. Hence

$$\lambda_u/\phi_u - \lambda_v/\psi_v = \phi_{uu}/\phi_u{}^2 - \psi_{vv}/\psi_v{}^2. \tag{6}$$

Therefore it remains to determine the conditions under which a function λ can satisfy an equation of this form.

Obtaining the cross derivative with respect to u and v and using this equation again, we find by (3) that, since the gaussian curvature G is not constant, there exists a function $\rho(u, v)$ such that

$$\phi_u = \rho G_u, \qquad \psi_v = \rho G_v. \tag{7}$$

Since ϕ depends upon u only and ψ depends upon v only, the function ρ must satisfy the conditions

$$(\partial/\partial u) \log \rho = -(\partial/\partial u) \log G_v, \qquad (\partial/\partial v) \log \rho = -(\partial/\partial v) \log G_u. \tag{8}$$

The compatibility condition for ρ yields the equation

$$(\partial^2/\partial u \partial v) \log G_u/G_v = 0. \tag{9}$$

This means that $\log G_u/G_v$ is a harmonic function, or that the gaussian curvature G is a function of a harmonic function.

Next substituting (7) into (6), we discover that

$$\lambda_u/G_u - \lambda_v/G_v = G_{uu}/G_u{}^2 - G_{vv}/G_v{}^2. \tag{10}$$

By equations (9) and (10), we have obtained the following analytic form of Kasner's theorem characterizing surfaces Σ applicable upon surfaces of revolution of variable gaussian curvature.

THEOREM 1. *A surface Σ with linear-element (1) is vase-like if, and only if, the system of two partial differential equations of the fourth order*

$$\begin{aligned} G_{uuv}/G_u - G_{uvv}/G_v &= G_{uv}(G_{uu}/G_u{}^2 - G_{vv}/G_v{}^2), \\ \lambda_u/G_u - \lambda_v/G_v &= G_{uu}/G_u{}^2 - G_{vv}/G_v{}^2, \end{aligned} \tag{11}$$

is identically satisfied.

Geometrically this result means that the ∞^1 curves $v' = G_u/G_v$ are an isothermal system of geodesics, as may be verified by substituting into the differential equation (4) of geodesics, and into the condition for isothermal

families. This isothermal system of geodesics of Σ will correspond to the meridians of the isometric surface of revolution. The parallels correspond to the curves $G = \text{const.}$, and also form an isothermal system.

THEOREM 2. *In any vase-like surface Σ which is conformally mapped upon a plane π, if the scale curves $\lambda = \text{const.}$, coincide with the family $G = \text{const.}$, then the scale curves are parallel straight lines or concentric circles in π. If $G = \text{const.}$, form a parallel family in the plane π, they coincide with the scale curves $\lambda = \text{const.}$, and hence are parallel straight lines or concentric circles.*

This is a consequence of the conditions (11) for a vase-like surface Σ and some theorems developed in one of our preceding papers.[2]

4. Vase-like surfaces Σ for which conformal maps exist such that the scale curves form isothermal systems which are neither parallel straight lines nor concentric circles. The reasons why we restrict ourselves to vase-like surfaces are the following. Let Σ be any surface applicable upon a surface of revolution. Then Σ is either developable, or sphere-like, or vase-like.

If Σ is developable, then in any conformal map of Σ upon a plane π, the scale curves form an isothermal family. Conversely any isothermal system of curves can serve as the scale curves of a conformal map of some developable surface Σ upon a plane π.

We have proved elsewhere that if a sphere-like surface Σ is conformally represented upon a plane π such that the scale curves form an isothermal family, then the scale curves must be either parallel straight lines or concentric circles.

The preceding statements demonstrate the reasons why we must restrict ourselves to vase-like surfaces. Also the isothermal systems of scale curves must be neither parallel straight lines nor concentric circles, for it is well-known that any surface Σ applicable upon a surface of revolution may be conformally mapped upon a plane π such that the scale curves are either parallel straight lines or concentric circles.

In the remainder of the paper, we shall prove the following result.

FUNDAMENTAL THEOREM 3. *The linear-elements of all the vase-like sur-*

---

[2] Kasner and De Cicco, "Scale curves in cartography," *Science*, vol. 98 (1943), pp. 324-325. Also "Scale curves in conformal maps," *Proceedings of the National Academy of Sciences*, vol. 30 (1944), pp. 162-164, and "Scale curves in general cartography," *Proceedings of the National Academy of Sciences*, vol. 30 (1944), pp. 211-215. Finally see "Geometry of scale curves in conformal maps," *American Journal of Mathematics*, vol. 67 (1945), pp. 157-166.

faces for which conformal maps exist such that the scale curves form isothermal systems which are neither parallel straight lines nor concentric circles can be reduced to the following types by appropriate conformal transformations

(12. 1) $$ds^2 = \frac{ce^{uv}}{(uv)^n}\,du\,dv,$$

(12. 2) $$ds^2 = \frac{2e^{f(u)+g(v)}}{(u+v)^2}\,du\,dv,$$

(12. 3) $$ds^2 = \frac{[(au+b)(cv+d)]^{n-2}}{(u+v)^n}\,du\,dv, \text{ where } n \neq 0, 2,$$

(12. 4) $$ds^2 = \frac{ce^{u+v}}{(u+v)^n}\,du\,dv, \text{ where } n \neq 0, 2,$$

(12. 5) $$ds^2 = \frac{c(uv)^{(n-2)/2}}{(u+v)^n}\,du\,dv, \text{ where } n \neq 0, 2,$$

(12. 6) $$ds^2 = \frac{c(uv)^k}{(uv-1)^n}\,du\,dv, \text{ where } k \neq 0,\ n-2, \text{ and } n \neq 0, 2,$$

(12. 7) $$ds^2 = \frac{c[u^{1+ib}v^{1-ib}]^{(n-2)/2}}{(u+v)^n}\,du\,dv, \text{ where } b \neq 0, \text{ and } n \neq 0, 2,$$

where $c \neq 0$, n, a, b, d, k, *are constants.*

In the real domain, there are essentially seven types of such vase-like surfaces, but in the imaginary domain there are only six types; the last two types, namely (12. 6) and (12. 7), can be made equivalent under an imaginary conformal transformation.

5. The beginning of the proof of our Fundamental Theorem 3. Any surface Σ, not necessarily vase-like, for which a conformal map exists such that the scale curves form an isothermal system has the linear-element of the form

(13) $$ds^2 = 2e^{\lambda[\alpha(U)+\beta(V)]}dU\,dV,$$

where $\alpha_U \neq 0$ and $\beta_V \neq 0$. Let $u = \alpha(U)$, $v = \beta(V)$. Noting that this transformation may be defined by $\log dU/du = \phi(u)$, $\log dV/dv = \psi(v)$, it is seen that the linear-element (13) may be written in the form

(14) $$ds^2 = 2e^{\lambda(w)+\phi(u)+\psi(v)}du\,dv,$$

where $w = u + v$. This is the form of the linear-element of that class of surfaces Σ for which there exists a conformal map of Σ upon a plane π such that the scale curves form an isothermal system.

The gaussian curvature of the surface Σ with the linear-element (14) is $G = -e^{-\lambda-\phi-\psi}\lambda_{ww}$. The partial derivatives of the first order with respect to u and v of G are

(15) $\quad G_u = -e^{-\lambda-\phi-\psi}\lambda_{ww}(g-\phi_u), \qquad G_v = -e^{-\lambda-\phi-\psi}\lambda_{ww}(g-\psi_v),$

where

(16) $$g = g(w) = \lambda_{www}/\lambda_{ww} - \lambda_w.$$

Since $G_u \neq 0$ and $G_v \neq 0$, we see that $\lambda_{ww} \neq 0$, $g \neq \phi_u$ and $g \neq \psi_v$. We omit the case where $\phi_u = \psi_v$, because otherwise we have the cases where the scale curves are parallel straight lines or concentric circles.

The condition that the curves $G = \text{const.}$, be an isothermal family reduces to

(17) $$(\partial^2/\partial u\partial v)\log(G_u/G_v) = (\partial^2/\partial u\partial v)\log(g-\phi_u)/(g-\psi_v) = 0.$$

The other condition that the orthogonal trajectories of $G = \text{const.}$, be a system of geodesics becomes

(18) $$\begin{aligned}(\lambda_w + g)/(g-\phi_u) - (\lambda_w + g)/(g-\psi_v) \\ = (g_w - \phi_{uu})/(g-\phi_u)^2 - (g_w - \psi_{vv})/(g-\psi_v)^2.\end{aligned}$$

Upon expanding (17) and making use of (18), we discover, since $\phi_u \neq \psi_v$, that

(19) $$g_{ww} = g_w\lambda_{www}/\lambda_{ww}.$$

From this, it is seen that there exist constants (a, b) such that

(20) $$g = (2a-1)\lambda_w + 2b.$$

Substituting this into (16), we find

(21) $$\lambda_{ww} = a\lambda_w^2 + 2b\lambda_w + c,$$

where (a, b, c) are constants, not all zero. This condition is only necessary for the solution of our problem.

If $a = 0$, we shall prove that the linear-element can be reduced to the form

(22) $$(\text{I}): \quad ds^2 = 2e^{uv+f(u)+g(v)}du\,dv.$$

If $a = b = 0$, then $c \neq 0$ since $\lambda_{ww} \neq 0$. Thus λ is a quadratic function of $w = u + v$. By an appropriate similitude in π, the linear-element (14) can be reduced to the form (I).

If $a = 0$ and $b \neq 0$, it is found by (21) that λ is given by an equation of the form

(23) $$\lambda = -\frac{cw}{2b} + h^2e^{2bw} + k,$$

where $h \neq 0$ and k are constants. The transformation $U = he^{2bu}$, $V = he^{2bv}$, reduces this to the form (I).

If $a \neq 0$, we shall prove that the linear-element can be reduced to the form

$$(24) \qquad \text{(II)}: \quad ds^2 = 2\,\frac{e^{f(u)+g(v)}}{(u+v)^n}\,du\,dv,$$

where n is a non-zero constant.

If $a \neq 0$ and $b^2 - ac = 0$, we find that λ is given by

$$(25) \qquad \lambda = -(b/a)(w - w_1) - (1/a)\log(w - w_0),$$

where (w_0, w_1) are constants. Let $n = 1/a$ and let $U = u - u_0$, $V = v - v_0$. Hence the linear-element (14) in this case is reduced to the form (II).

Finally let $a \neq 0$ and $b^2 - ac \neq 0$. The complete solution of (21) is

$$(26) \qquad \lambda = -(b/a)(w - w_1) - (1/a)\log\left[e^{d(w-w_0)} - e^{-d(w-w_0)}\right],$$

where $d^2 = b^2 - ac$, w_0, w_1, are constants. Hence the linear-element (14) in this case may be written as

$$(27) \qquad ds^2 = \frac{2e^{f(u)+g(v)}}{\left[e^{2d(w-w_0)} - 1\right]^n}\,du\,dv,$$

where $n = 1/a \neq 0$. Now let

$$(28) \quad (2U - 1)/(2U - 1) = e^{2d(v-v_0)}, \quad (2V + 1)/(2V - 1) = e^{2d(u-u_0)}.$$

(If d is pure imaginary, replace the ± 1 in the first equation by $\pm i$ and the ± 1 in the second equation by $\pm i$). Substituting (28) into (27), we see that the linear-element (14) is changed to the form (II).

6. The discussion of the Case (I) where the linear-element of the surface Σ is given by (22). The gaussian curvature is $G = -e^{-uv-f-g}$. The condition that $G =$ const., be an isothermal family becomes

$$(29) \qquad f_{uu}/(v + f_u)^2 = g_{vv}/(u + g_v)^2.$$

The other condition reduces to this also.

If f_{uu} or g_{vv} is zero, then both are zero. Hence f and g are both linear functions in u and v separately. Under an appropriate translation, we find that case (I) in this instance can be reduced to (12.1), where $n = 0$, of our Fundamental Theorem 3.

If f_{uu} or g_{vv} is not zero, then both are not zero. The cross derivative of the square root of the reciprocal of (29) is

$$(30) \qquad (\partial/\partial u)\,(f_{uu})^{-1/2} = \pm\,(\partial/\partial v)\,(g_{vv})^{-1/2}.$$

This demonstrates that

$$(31)\qquad f_u = -a^2/(u-u_0)+b, \qquad g_v = -a^2/(v-v_0)+c,$$

where $(a \neq 0, b, c)$ are constants. The substitution of (31) into (29) shows that $c = -u_0$, $b = -v_0$. Hence f and g are of the forms

$$(32)\quad f = -a^2 \log(u-u_0) - v_0 u + u_1, \quad g = -a^2 \log(v-v_0) - u_0 v + v_1,$$

where $(a \neq 0, u_0, v_0, u_1, v_1)$ are constants. Under an appropriate translation, our linear-element can be reduced to the form (12.1) of our Fundamental Theorem 3. In the real domain, n is a positive real constant.

7. The discussion of the Case (II) where the linear-element of the surface Σ is given by (24). The gaussian curvature is $G = -n(u+v)^{n-2} \times e^{-f(u)-g(v)}$. Note that $n \neq 0$. The condition that $G =$ const., be an isothermal family is

$$(33)\qquad \frac{(n-2)f_{uu}+f_u^2}{[(u+v)f_u-(n-2)]^2} = \frac{(n-2)g_{vv}+g_v^2}{[(u+v)g_v-(n-2)]^2}.$$

The other condition that the orthogonal trajectories of $G =$ const., form a geodesic system may be shown to be equivalent to this.

If $n = 2$, the above equation is an identity yielding the linear-element (12.2) of our Fundamental Theorem 3.

Note that $n = 0$ yields a developable surface, which case is excluded from consideration. Henceforth $n \neq 0, 2$.

Let $f_u = g_v = 0$. The linear-element in this case assumes the form (12.3), where $a = c = 0$.

In the imaginary domain, we may have the case where $f_u = 0$ but $g_v \neq 0$; or $f_u \neq 0$ but $g_v = 0$. Here we have the linear-element (12.3), where $a = 0$ or $c = 0$.

If $n \neq 0, 2$, and if either of the numerators of (33) is zero but $f_u \neq 0$ and $g_v \neq 0$, it follows that in this case the linear-element is of the form (12.3) where $a \neq 0$ and $c \neq 0$.

Finally we consider the case where either numerator of (33) is not zero. Then both numerators are not zero. Then the cross derivative of the square root of the reciprocal of (33), is

$$(34)\qquad \frac{\partial}{\partial u}\,\frac{f_u}{[(n-2)f_{uu}+f_u^2]^{1/2}} = \pm\frac{\partial}{\partial v}\,\frac{g_v}{[(n-2)g_v+g_v^2]^{1/2}}.$$

Hence each side represents the same constant.

First we must consider the case when this constant is zero. We find

$$(35)\qquad (n-2)f_{uu}/f_u^2 = a^2-1, \qquad (n-2)g_{vv}/g_v^2 = b^2-1,$$

where a and b are constants. Substituting (35) into (33), we find

$$\frac{a^2 f_u^2}{[(u+v)f_u-(n-2)]^2}=\frac{b^2 g_v^2}{[(u+v)g_v-(n-2)]^2}. \tag{36}$$

Since we are working under the condition where neither numerator of (33) is zero, we see that $a\neq 0$, $b\neq 0$, $f_u\neq 0$, $g_v\neq 0$. Taking the cross derivative of the reciprocal of (36) and making use of (36) we find that $a^2=b^2$. From (36), we obtain

(37) Either $f_u=g_v$, or $(n-2)/f_u+(n-2)/g_v-2(u+v)=0$.

If $n\neq 2$, $f_u\neq 0$, $g_v\neq 0$ and $f_u=g_v$, it is found that under an appropriate similitude the linear-element of Σ can be reduced to the form (12.4) of our Fundamental Theorem 3.

Next if the second of the conditions (37) is valid, it is found, by (35), that $b^2=a^2=-1$. Thus we find by (35) and the second of the conditions (37) that

$$\begin{aligned} f&=[(n-2)/2]\log(u-u_0)+\text{const.},\\ g&=[(n-2)/2]\log(v+u_0)+\text{const.} \end{aligned} \tag{38}$$

By an appropriate translation, we find that if $n\neq 0, 2$, and if the equation (35) together with the second of the conditions (37), hold, the linear-element can be reduced to the form (12.5).

Finally we have to consider the case where the equal constant of (34) is not zero. Hence we find

$$\frac{(n-2)f_{uu}+f_u^2}{f_u^2}=\frac{a^2}{(u-u_0)^2},\quad \frac{(n-2)g_{vv}+g_v^2}{g_v^2}=\frac{a^2}{(v-v_0)^2}, \tag{39}$$

where $(a\neq 0, u_0, v_0)$ are constants. Substitute this into (33) and we find

$$\frac{f_u^2}{(u-u_0)^2[(u+v)f_u-(n-2)]^2}=\frac{g_v^2}{(v-v_0)^2[(u+v)g_v-(n-2)]^2}. \tag{40}$$

From this equation, it follows that

$$(u-u_0)[(u+v)-(n-2)/f_u]=\pm(v-v_0)[(u+v)-(n-2)/g_v]. \tag{41}$$

Differentiating this equation with respect to v and also with respect to u and making use of (39), we discover that the minus sign is impossible and we get

$$f_u=\frac{(n-2)(u-u_0)}{(u+v_0)(u-u_0)+a^2},\quad g_v=\frac{(n-2)(v-v_0)}{(v+u_0)(v-v_0)+a^2}. \tag{42}$$

The equation (33) is obviously satisfied by these.

10

For this general case, let us apply the translation $U = u + \frac{1}{2}(-u_0 + v_0)$, $V = v + \frac{1}{2}(u_0 - v_0)$. Of course, the linear-element of Σ is of the same form (II) but f and g are now given by

$$f_u = \frac{(n-2)(u-\alpha)}{u^2 - b^2}, \qquad g_v = \frac{(n-2)(v-\alpha)}{v^2 - b^2}, \tag{43}$$

where $\alpha = \frac{1}{2}(u_0 + v_0)$ and $b^2 = \alpha^2 - a^2$. Since $a \neq 0$, it follows that b cannot be equal to $\pm\alpha$.

If $b = 0$, then $\alpha \neq 0$, and we find that f and g are given by

$$f = (n-2)(\log u + \alpha/u) + \text{const.}, \quad g = (n-2)(\log v + \alpha/v) + \text{const.} \tag{44}$$

Substituting these into Case (II) and then using the substitution $U = 1/u$, $V = 1/v$, we see that our linear-element is reduced to the form (12.4) of our Fundamental Theorem 3.

Finally if $b \neq 0$, then f and g are given by

$$\begin{aligned} f &= [(n-2)/2b][(b-\alpha)\log(u-b) + (b+\alpha)\log(u+b)] + \text{const.} \\ g &= [(n-2)/2b][(b-\alpha)\log(v-b) + (b+\alpha)\log(v+b)] + \text{const.} \end{aligned} \tag{45}$$

It follows that the linear-element of Σ is of the form

$$\begin{aligned} ds^2 = [c/(u+v)^n][\{(u-b)(v-b)\}^{(1-\alpha/b)} \\ \times \{(u+b)(v+b)\}^{(1+\alpha/b)}]^{(n-2)/2} du\, dv, \end{aligned} \tag{46}$$

where $a \neq 0$, $b \neq 0$, and $\alpha \neq \pm b$.

If b is real, the transformation:

$$u = b(U-1)/(U+1), \qquad v = b(V-1)/(V+1), \tag{47}$$

will change the form of (46) into the form (12.6).

On the other hand if b is pure imaginary, the transformataion:

$$= b(U-1)/(U+1), \qquad v = -b(V-1)/(V+1), \tag{48}$$

will change the form (46) into the form (12.7).

This completes the proof of our Fundamental Theorem 3.

8. Conclusion. It is remarked that the conformal maps with isothermal systems of scale curves of the vase-like surfaces of our Fundamental Theorem 3, can be obtained by compounding inversely the various correspondences which were used in order to reduce the linear-element (13) to these linear-elements (12).

ILLINOIS INSTITUTE OF TECHNOLOGY,
CHICAGO, ILLINOIS.

COMPLEX FUNCTIONS POSSESSING DIFFERENTIALS.*

By VINCENT C. POOR.

I. THE DIFFERENTIALS OF FUNCTIONS.

1. Introduction. We shall be interested in the complex function $f(z, z')$ of two complex variables z and z'. We shall also consider incidentally the polygenic function $f(z)$ obtained by replacing z' by $\bar{z}$ the conjugate of z. It is polygenic in the sense that its z derivative at a point depends on the direction of approach to the point.

2. Purpose and content. The purpose of this paper is to study differentials of complex functions, in particular the restricted Hamilton [1] and the Young [2] differentials and their relations to each other. Closely related to the restricted Hamilton differential are the Stolz [3] and the Rainich [4] differentials; the latter is a Hamilton differential restricted by the linearity property and in addition a property something less than continuity.

The Young definition is also given by Mrs. Young [5] in discussing functions possessing differentials. In her paper necessary and sufficient conditions for the existence of a Young differential of a complex function are given. However, some of these conditions, to say the least, are superfluous, and properly formulated the theorem is not proved.

The necessary and sufficient conditions for the existence of a Young differential will be set up here and proved. Also certain results of the Fundamenta paper will be generalized to the complex plane. Finally, a power-series expansion of a polygenic function will be given.

3. Regularity. The Hamilton differential was invented to make the absolute geometric development of vector analysis possible; the differential quotient or derivative of a point function is non-existent since division by a

* Received March 23, 1945.

[1] V. C. Poor, "On the Hamilton differential," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 945-948.

[2] W. H. Young, "On differentials," *Proceedings of the London Mathematical Society*, Series 2, vol. 7 (1908), p. 162.

[3] O. Stolz, *Differential und integral Rechnung*, vol. I, p. 132.

[4] G. Y. Rainich, *American Journal of Mathematics*, vol. 46 (1924), p. 78.

[5] Grace Chisholm Young, "On functions possessing differentials," *Fundamenta Mathematicae*, Tome 15 (1930), pp. 61-94.

vector is excluded. This definition will be restricted by a linearity condition making it essentially equivalent to the Stolz differential. For brevity we shall call a function regular if it possesses a differential, R-regular if it possesses a restricted Hamilton differential and Y-regular if it possesses a Young differential.

4. The restricted Hamilton differential. The Hamilton differential may be expressed by the equation

$$(4.1\text{ a})\qquad f'(P_1, dP) = \lim_{\lambda \to 0} \frac{f(P_1 + \lambda dP) - f(P_1)}{\lambda}$$

where P_1 is a point of the two-dimensional vector space z, z', while dP is an arbitrary displacement of P_1. The restricted Hamilton differential is a Hamilton differential (4. 1 a) which satisfies the requirement that it be linear. The linearity property is given by

$$(4.1\text{ b})\qquad f'(P_1, dz\mathbf{i} + dz'\mathbf{j}) = f'(P_1, \mathbf{i})\,dz + f'(P_1, \mathbf{j})\,dz'$$

where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors in the complex vector space.

THEOREM 1. *The necessary and sufficient condition for the complex function $f(z, z')$ to be R-regular at a point $P_1 \equiv (z_1, z'_1)$ is that $f(z, z')$ be expressible in the form*

$$(4.2)\qquad f(z, z') = f(z_1, z'_1) + \frac{\partial f}{\partial z_1}(z - z_1) + \frac{\partial f}{\partial z'_1}(z' - z'_1) + \eta_1$$

where η_1 is an infinitesimal function of $z - z_1$ and $z' - z'_1$ of an order higher than the first.

In proving the necessity of the condition one observes that

$$f(P_1 + \lambda dP) = f(P_1) + \lambda f'(P_1, dP) + \eta_1$$

follows from (4. 1 a), while the linearity condition (4. 1 b) changes this equation into

$$(4.3)\qquad f(P_1 + \lambda dP) = f(P_1) + f'(P_1, \mathbf{i})\lambda dz + f'(P_1, \mathbf{j})\lambda dz' + \eta_1$$

where η_1 is an infinitesimal function of λ of order higher than the first. Evidently (4. 3) may be put into the coordinate form

$$f(z_1 + \lambda dz, z'_1 + \lambda dz') = f(z_1, z'_1) + f'(P_1, \mathbf{i})\lambda dz + f'(P_1, \mathbf{j})\lambda dz' + \eta_1$$

and when $z - z_1$ and $z' - z'_1$ replace λdz and $\lambda dz'$ respectively, this last equation becomes

$$f(z, z') = f(z_1, z'_1) + f'(P_1, \mathbf{i})(z - z_1) + f'(P_1, \mathbf{j})(z' - z'_1) + \eta_1$$

where η_1 is now an infinitesimal function of $z - z_1$ and $z' - z'_1$ of higher order than the first. That $f'(P_1, \mathfrak{i})$ and $f'(P_1, \mathfrak{j})$ are the partial derivatives is shown by expanding the right member of (4. 1 a), when expressed in coördinate form. Thus,

$$\begin{aligned}
&\lim_{\lambda\to 0} \frac{f(z_1 + \lambda dz, z'_1 + \lambda dz') - f(z_1, z'_1)}{\lambda} \\
&\quad = \lim_{\lambda\to 0} \frac{f(z_1 + \lambda dz, z'_1 + \lambda dz') - f(z_1, z'_1 + \lambda dz')}{\lambda dz} dz \\
&\quad + \frac{f(z_1, z'_1 + \lambda dz') - f(z_1, z'_1)}{\lambda dz'} dz \\
&\quad = \frac{\partial f}{\partial z} dz + \frac{\partial f}{\partial z'} dz'.
\end{aligned}$$

and when this is identified with the right member of (4. 1 b) the statement is proved and also the necessity of the condition.

The proof of the sufficiency is very simple. Our hypothesis is given by (4. 2). Using the results obtained (4. 3) follows when $z - z_1$ and $z' - z'_1$ are replaced by λdz and $\lambda dz'$ respectively in (4. 2), making η_1 an infinitesimal function of λ of higher order than the first. Thus when $f(P_1)$ is transposed and a division by λ made in (4. 3), one finds in the limit that

$$f'(P_1, dP) = f'(P_1, \mathfrak{i})\, dz + f'(P_1, \mathfrak{j})\, dz'$$

which shows that the Hamilton definition (4. 1 a) is satisfied and also that the linearity restriction (4. 1 b) holds.

Another set of necessary and sufficient conditions for R-regularity is the existence and finiteness of the partial derivatives $\partial f/\partial z_1$, $\partial f/\partial z'_1$. This is a consequence of Theorem 2. However, Theorem 1, which implies (4. 2), also implies the existence and finiteness of these partial derivatives. The conditions are thus necessary.

For the sufficiency proof one can construct (4. 2) from the hypothesis, since $\partial f/\partial z_1$ and $\partial f/\partial z'_1$ imply the continuity of $f(z, z'_1)$ and $f(z_1, z')$ at (z_1, z'_1) and thus the existence and finiteness of f at (z_1, z'_1). But (4. 2) implies R-regularity. It may be observed that $f(z, z'_1)$ and $f(z_1, z')$ are each analytic at (z_1, z'_1). Incidentally the theorem makes the restricted Hamilton and the Stolz differentials essentially equivalent.

5. The Young differential. This differential is given in the Fundamenta paper as follows:

If

$$(5.1) \qquad \begin{aligned} f(z_1 + \lambda dz, z'_1 + \mu dz') &- f(z_1, z'_1) \\ &= [A + e_1(\lambda, \mu)]\lambda dz + [B + e_2(\lambda, \mu)]\mu dz' \end{aligned}$$

where A and B are constants depending on the values of z_1 and z'_1 and λ and μ are arbitrary real scalar parameters, while $e_1(\lambda,\mu)$ and $e_2(\lambda,\mu)$ functions of z and z' are infinitesimal functions of λ and μ, going to zero when λ and μ are made to go to zero in any way whatever; then $f(z,z')$ is said to possess a differential at $P_1 = (z_1, z'_1)$; which may be expressed in the form $Adz + Bdz'$.

In this form the factors λ and μ are thought of as absorbed by dz and dz' respectively since dz and dz' are just as arbitrary without these factors. Also, if, in the complex function $f(z,z')$, z and z' are regarded as real variables the definition just given is exactly the one given in the Fundamenta paper aside from a slight change of notation i. e. λdz and $\mu dz'$ replace h and k respectively. Thus the complex functions, treated there, are included in $f(z,z')$.

THEOREM 2. *For the complex function $f(z,z')$ to be Y-regular at $P_1 = (z_1, z'_1)$ it is necessary and sufficient that*

(5. 2 a) $\quad \partial f/\partial z_1$ *and* $\partial f/\partial z'_1$ *exist and are finite, and*

(5. 2 b) $$\lim_{\lambda,\mu\to 0} \phi(\lambda,\mu)/\lambda = 0; \qquad \lim_{\lambda,\mu\to 0} \phi(\lambda,\mu)/\mu = 0$$

where by definition

$$\phi(\lambda,\mu) \equiv f(z_1+\lambda dz, z'_1+\mu dz') - f(z_1, z'_1+\mu dz') - f(z_1+\lambda dz, z'_1) + f(z_1, z'_1) \tag{5. 3}$$

To prove the necessity of the primary conditions (5. 2 a) we let μ go to zero first in (5. 1); this makes

$$f(z_1+\lambda dz, z'_1) - (z_1, z'_1) = [A + e_1(\lambda, 0)]\lambda dz.$$

But

$$\lim_{\lambda\to 0} \frac{f(z_1+\lambda dz, z'_1) - f(z_1, z'_1)}{\lambda dz} = \lim_{\lambda\to 0} A + e_1(\lambda, 0)$$

or

$$\partial f/\partial z_1 = A$$

since $e_1(\lambda, 0)$ goes to zero with λ by hypothesis. In a similar way we find by letting λ go to zero first in (5. 1) that

$$\partial f/\partial z'_1 = B.$$

To prove the necessity of the secondary conditions (5. 2 b) we use the results just obtained to exhibit explicitly e_1 and e_2 for any complex function. In fact (5. 1) may be written in the form

$$(5.4)\quad f(z_1 + \lambda dz, z'_1 + \mu dz') - f(z_1, z'_1)$$

$$\equiv \phi(\lambda, \mu) + f(z_1 + \lambda dz, z'_1) - f(z_1, z'_1) + f(z_1, z'_1 + \mu dz') - f(z_1, z'_1)$$

$$\equiv \left[\tau \frac{\phi(\lambda, \mu)}{\lambda dz} + A - \frac{\partial f}{\partial z_1} + \frac{f(z_1 + \lambda dz, z'_1) - f(z_1, z'_1)}{\lambda dz}\right] \lambda dz$$

$$+ \left[\sigma \frac{\phi(\lambda, \mu)}{\mu dz'} + B - \frac{\partial f}{\partial z'_1} + \frac{f(z_1, z'_1 + \mu dz') - (z_1, z'_1)}{\mu dz'}\right] \mu dz',$$

where $\tau + \sigma = 1$. From this last form we read off e_1 and e_2:

$$(5.5)\quad \begin{aligned} e_1(\lambda, \mu) &\equiv \tau \frac{\phi(\lambda, \mu)}{\lambda dz} - \frac{\partial f}{\partial z_1} + \frac{f(z_1 + \lambda dz, z'_1) - f(z_1, z'_1)}{\lambda dz} \\ e_2(\lambda, \mu) &\equiv \sigma \frac{\phi(\lambda, \mu)}{\mu dz'} - \frac{\partial f}{\partial z'_1} + \frac{f(z_1, z'_1 + \mu dz') - f(z_1, z'_1)}{\mu dz'}. \end{aligned}$$

By hypothesis e_1 and e_2 go to zero when λ and μ are made to go to zero in any manner whatever and since the second and third terms of e_1 and e_2 cancel in the limit it follows that

$$(5.6)\quad \lim_{\lambda, \mu \to 0} \tau\phi(\lambda, \mu)/\lambda = \lim_{\lambda, \mu \to 0} \sigma\phi(\lambda, \mu)/\mu = 0.$$

Since this is true for every choice of τ and σ one concludes that (5.2 b) holds. This proves the necessity of the condition.

Suppose that as $\lambda, \mu \to 0$, $\tau \to 0$ and $\sigma \to 1$; this would make $\phi(\lambda, \mu)/\mu \to 0$ while an interchange of τ and σ makes $\phi(\lambda, \mu)/\lambda \to 0$. The disturbing case is when $\tau \to \infty$, $\sigma \to -\infty$ say, while $\tau + \sigma \to 1$ with $\lambda, \mu \to 0$. Since (5.6) has to be satisfied in this case it follows that $1/\tau$ and $1/\sigma$ are infinitesimal functions of λ and μ of a lower order than $\phi(\lambda, \mu)/\lambda$ and $\phi(\lambda, \mu)/\mu$ respectively. This result will be used in the sufficiency proof.

In the sufficiency proof we may take

$$\partial f/\partial z_1 = A \quad \text{and} \quad \partial f/\partial z'_1 = B$$

since by hypothesis these partial derivatives exist and are finite. This permits us to write (5.1) in form (5.4) from which we may again read off e_1 and e_2 as given in (5.5). The hypothesis (5.2 b) together with the restrictions imposed on τ and σ in the last paragraph, evidently imply the validity of (5.6). This makes the first term in each of e_1 and e_2 zero in the limit while the second terms in each cancel in the limit as before. From the hypotheses we have thus constructed (5.1) wherein e_1 and e_2 have the prescribed properties.

6. Thomae's test. The usual formulation of this test is as follows: The existence and finiteness of $\partial f/\partial z_1$, $\partial f/\partial z'_1$ together with the continuity of one, say $\partial x/\partial z$, implies regularity. Evidently continuity of a partial derivative is unnecessary for R-regularity; while continuity of just one of the existing partial derivatives is not sufficient for Y-regularity. For if it were, the secondary conditions would be satisfied. We find that

$$\lim_{\lambda,\mu\to 0}\frac{\phi(\lambda,\mu)}{\lambda dz}=\lim_{\lambda,\mu\to 0}\frac{f(z_1+\lambda dz, z'_1+\mu dz')-f(z_1, z'_1+\mu dz')}{\lambda dz}$$
$$-\frac{f(z_1+\lambda dz, z'_1)-f(z_1, z'_1)}{\lambda dz}$$
$$=\lim_{\lambda\to 0}\frac{\partial f(z_1, z'_1+\mu dz')}{\partial z_1}-\frac{\partial f}{\partial z_1}$$

which vanishes only if $\partial f/\partial z_1$ is continuous in z'. In a similar way one finds that

$$\lim_{\lambda,\mu\to 0}\frac{\phi(\lambda,\mu)}{\mu dz'}=\lim_{\lambda\to 0}\frac{\partial f(z_1+\lambda dz, z'_1)}{\partial z'_1}-\frac{\partial f}{\partial z'_1}$$

vanishes if $\partial f/\partial z'_1$ is continuous in z.

Continuity of one partial derivative seems to be of little value as a test for Y-regularity. The continuity of one must be supplemented by some other condition possibly by the continuity of the other.

7. *R*-*Y*-regularity relations. It should be expected that Y-regularity implies R-regularity. That the latter is a special case of the former is seen by replacing μ by λ in $\phi(\lambda,\mu)$ as defined in (5. 3). We find that

$$\frac{\phi(\lambda,\mu)}{\lambda dz}=\frac{f(z_1+\lambda dz, z'_1+\lambda dz')-f(z_1, z'_1+\lambda dz')}{\lambda dz} \tag{7.1}$$
$$-\frac{f(z_1+\lambda dz, z'_1)-f(z_1, z'_1)}{\lambda dz}$$

which becomes zero in the limit. Thus the secondary conditions (5. 2 b) are seen to be identically satisfied.

The question naturally arises: What less restrictive condition on the definition for Y-regularity is sufficient for an R-regular function to be at the same time a Y-regular function. A sufficient restriction is that λ and μ be infinitesimals of the same order. In fact when $k\lambda$ is substituted for μ in $\phi(\lambda,\mu)$ (7. 1) will remain unchanged, since dz' will merely carry the additional factor k which may be absorbed by it.

By imposing a restriction on the function rather than on the definition for Y-regularity we have

THEOREM 3. *For an R-regular function $f(z, z')$ to be at the same time a Y-regular function it is sufficient for $\partial f/\partial z$ to be continuous in z' and $\partial f/\partial z'$ to be continuous in z at P_1.*

That this theorem is true follows from the discussion of the Thomae test. In fact it was seen there that when the conditions of the theorem are applied to $f(z, z')$ the secondary conditions (5.2 b) are identically satisfied.

8. The geometry. The special class of polygenic functions considered here are obtained from the function $f(z, z')$ by making $z' = \bar{z}$ the conjugate of z. We first superimpose the z' complex plane on the z plane making the origins and corresponding axes incident. If we now replace z' by $\bar{z}$ in $f(z, z')$ this function becomes the polygenic function $f(z)$ in the sense that its z derivative at a point depends on the direction of approach to the point.

9. *R*-regularity of *f*(*z*). The replacement of z' by $\bar{z}$ changes (4.2) into

$$f(z) = f(z_1) + \frac{\partial f}{\partial z_1}(z - z_1) + \frac{\partial f}{\partial \bar{z}_1}(\bar{z} - \bar{z}_1) + \eta_1 \tag{9.1}$$

where $\partial f/\partial z_1$ and $\partial f/\partial \bar{z}_1$ are written for $\partial f(z_1, z'_1)/\partial z_1$ and $\partial f(z_1, z'_1)/\partial z'_1$ respectively wherein z'_1 is replaced by $\bar{z}_1$.

THEOREM 4. *The necessary and sufficient conditions that the polygenic function $f(z)$ be R-regular at a point are the existence and finiteness of the areal and mean derivatives at the point.*[6]

In the proof of this theorem we note that (4.2) furnished the necessary and sufficient condition for R-regularity of $f(z, z')$. Thus (9.1), a consequence of (4.2), furnishes the necessary and sufficient condition for R-regularity of $f(z)$. Also since the existence and finiteness of $\partial f/\partial z_1$ and $\partial f/\partial z'_1$ furnished a second set of necessary and sufficient conditions for R-regularity of $f(z, z')$ all that is left to do is to prove that $\partial f/\partial z_1$ and $\partial f/\partial \bar{z}_1$ are the areal and mean derivatives respectively.

By definition the areal derivative

$$\frac{\partial f}{\partial \alpha} = \operatorname*{limit}_{\sigma \to 0} \frac{1}{2i\sigma} \int_C f(z)\, dz = \operatorname*{limit}_{r \to 0} \frac{1}{2\pi r^2 i} \int_0 f(z)\, dz$$

[6] V. C. Poor, *Transactions of the American Mathematical Society*, vol. 32 (1930), p. 216.

where σ is chosen as the area of the circle radius r center at $z=z_0$ while the integral is to be taken around the circle. Briefly then we put $z-z_1=re^{i\theta}$ and replace $f(z)$ by the right member of (9.1); we find that

$$\frac{\partial f}{\partial \alpha}=\lim_{r\to 0}\frac{1}{2\pi r^2 i}\ \frac{\partial f}{\partial z_1}\int_0 (\bar z-\bar z_1)\,dz=\lim_{r\to 0}\frac{1}{2\pi r^2 i}\ \frac{\partial f}{\partial \bar z_1}\int_0 2\pi r^2 i\,d\theta=\frac{\partial x}{\partial \bar z_1}.$$

All other integrals vanish or become zero in the limit. In exactly the same way we show that

$$\frac{\partial f}{\partial \beta}=\lim_{r\to 0}\frac{-1}{2\pi i}\int_0 f(z)\,d\bar z=\frac{\partial f}{\partial z_1}.$$

10. Y-regularity of $f(z, z')$.

THEOREM 5. *The necessary and sufficient conditions for the existence of a Y-differential of the polygenic function $f(z)$ at $z=z_1$ are*

(a) *the existence and finiteness of the areal and mean derivatives*

$$\partial f/\partial\alpha \quad \textit{and} \quad \partial f/\partial\beta$$

and

(b) $$\lim_{\lambda,\mu\to 0}\frac{\phi(\lambda,\mu)}{\lambda}=0=\lim_{\lambda,\mu\to 0}\frac{\phi(\lambda,\mu)}{\mu}.$$

This should be evident from the geometry. However, there is a lack of symmetry due to the μ. Putting $z-z_1=\lambda dz$ and therefore $\bar z-\bar z_1=\lambda d\bar z$ into $f(z_1+\lambda dz, \bar z_1+\mu d\bar z)$ we find that

$$f[\bar z, \bar z_1+(\mu/\lambda)(\bar z-\bar z_1)]$$
$$=f(z_1,\bar z_1)+[A+e_1(\lambda,\mu)](z-z_1)+[B+e_2(\lambda,\mu)](\mu/\lambda)(\bar z-\bar z_1).$$

This however may be used for $f(z)$ since as r approaches zero this approaches $f(z_1)$ as a limit. We may thus use this for $f(z)$ in determining the areal and mean derivatives so that $\partial f/\partial\alpha=B$ and $\partial f/\partial\beta=A$ as before.

Also $e_1(\lambda,\mu)$ and $e_2(\lambda,\mu)$ may be constructed as previously, noting that $\partial f/\partial\bar z_1$ and $\partial f/\partial z_1$ which are defined above are the areal and mean derivatives respectively. Further argument follows that in Theorem 2. This should be conclusive without further detail.

II. THE SUGGESTED GENERALIZATIONS.

11. The line integral. We have seen that the regular function $f(z, z')$ is analytic in z and z' in their separate domains. The line integral of this function from z_0 to z_1 is thus independent of the path of integration.

We define $F(z_1, z'_1)$ by the equation

$$F(z_1, z'_1) = \int_{z_0}^{z_1} f(z, z'_1)\,dz \tag{11.1}$$

where the integral is to be taken along a rectifiable path in the domain D.

THEOREM 6. *If $f(z, z')$ is Y-regular at (z_1, z'_1) its integral $F(z_1, z'_1)$ has a double incrementary ratio which tends to the limit $\partial f(z_1, z'_1)/\partial z'_1$ or $\partial f(z, z')/\partial z'$ at (z_1, z'_1) that is*

$$\begin{aligned}
&\operatorname*{limit}_{\lambda,\mu\to 0} \frac{\Phi(\lambda,\mu)}{\lambda dz \mu dz'} \\
&= \operatorname*{limit}_{\lambda,\mu\to 0} \frac{F(z_1+\lambda dz, z'_1+\mu dz') - F(z_1+\lambda dz, z'_1) - F(z_1, z'_1+\mu dz') + F(z_1, z'_1)}{\lambda dz \mu dz'} \\
&= \frac{\partial f(z_1, z'_1)}{\partial z'_1}.
\end{aligned}$$

Expressed in terms of the integral

$$\begin{aligned}
\Phi(\lambda,\mu) &= \int_{z_0}^{z_1+\lambda dz} f(z, z'_1+\mu dz')\,dz - \int_{z_0}^{z_1+\lambda dz} f(z, z'_1)\,dz \\
&\quad - \int_{z_0}^{z_1} [f(z, z'_1+\mu dz') - f(z, z'_1)]\,dz \\
&= \int_{z_1}^{z_1+\lambda dz} [f(z, z'_1+\mu dz') - f(z, z'_1)]\,dz.
\end{aligned} \tag{11.2}$$

If this integral is transformed by the substitution $z = z_1 + \lambda s$ one finds that

$$\begin{aligned}
&\operatorname*{limit}_{\lambda,\mu\to 0} \frac{\Phi(\lambda,\mu)}{\lambda dz \mu dz'} \\
&\quad = \operatorname*{limit}_{\lambda,\mu\to 0} \frac{1}{dz} \int_0^{dz} \frac{f(z_1+\lambda s, z'_1+\mu dz') - f(z_1+\lambda dz, z'_1)}{\mu dz'}\,ds \\
&\quad = \operatorname*{limit}_{\lambda,\mu\to 0} \frac{1}{dz} \int_0^{dz} \frac{\partial f(z_1, z'_1)}{\partial z'_1}\,ds = \frac{\partial f(z_1, z'_1)}{\partial z'_1}.
\end{aligned}$$

However if μ is made to go to zero first the limit is $\partial f(z_1 + \lambda s, z'_1)/\partial z'_1$ which becomes $\partial f(z_1, z'_1)/\partial z'_1$ if $\partial f/\partial z'_1$ is continuous at (z_1, z'_1).

We may observe, as a corollary, that if $f(z, z')$ is R-regular at $(z_1, z'_1$ the theorem is valid without the continuity condition since, in this case $\mu = k\lambda$. Also, since the above result is true at any point z_1, z'_1 in the domain D and its conjugate $D' = \bar{D}$, z'_1 may be replaced by $\bar{z}_1$ the conjugate to z. But we have seen that $\partial f(z_1, \bar{z}_1)/\partial \bar{z}_1$ is the areal derivative of $f(z)$ at $z = z_1$. The

result, of course, does not imply that the integrand in $F(z, z')$ is $f(z)$; the integral would in that case depend on the path. A similar result may be obtained by taking the path of integration from z'_0 to z'_1 in the conjugate domain D' with z' as the variable of integration, leading finally to the mean derivative in the limit.

THEOREM 7. *If the integral* $\int_{z_0}^{z_1} f(z, z'_1)\,dz$ *is independent of the path and if* $f(z, z')$ *is Y-regular not only at* (z_1, z'_1) *but also at every point of the path of integration from* z_0 *to* z_1 *then* $\partial f/\partial z'_1$ *is an integrable function of* z *and* $\partial F/\partial z'_1$ *exists and is given by differentiation under the integral sign.*

In the proof we write the previous theorem

$$\operatorname*{limit}_{\lambda,\mu\to 0} \frac{1}{\lambda dz \mu dz'} \int_{z_1}^{z_1+\lambda dz} [f(z, z'_1 + \mu dz') - f(z, z'_1)]\,dz = \frac{\partial f(z_1, z'_1)}{\partial z'_1}.$$

Hence

$$\operatorname*{limit}_{\lambda\to 0} \frac{1}{\lambda dz} \int_{z_1}^{z_1+\lambda dz} [f(z, z'_1 + \mu dz') - f(z, z'_1)]\,dz = \frac{\partial f(z, z'_1)}{\partial z'_1}\mu dz' + e(\mu)\mu dz'$$

where $e(\mu)$ is an infinitesimal function of μ. But according to Lebesgue's theorem [7] the left member is equal to the integrand at $z = z_1$. Thus

$$\frac{1}{\mu dz'}[f(z, z'_1 + \mu dz') - f(z, z'_1)] = \frac{\partial f(z, z'_1)}{\partial z'_1} + e(\mu)$$

where we have written z for z_1. When this result is multiplied by dz and integrated along the rectifiable path one finds that

$$\frac{1}{\mu dz'} \int_{z_0}^{z_1} [f(z, z'_1 + \mu dz' - f(z, z'_1)]\,dz = \int_{z_0}^{z_1} \frac{\partial f(z, z'_1)}{\partial z'_1}\,dz + \int_{z_0}^{z_1} e\,dz.$$

When the left member is identified through the definition (11.1) of $F(z_1, z'_1)$ we find that

$$\frac{F(z_1, z_1 + \mu dz') - F(z_1, z'_1)}{\mu dz'} = \int_{z_0}^{z_1} \frac{\partial f(z, z'_1)}{\partial z'_1}\,dz + \int_{z_0}^{z_1} e\,dz.$$

In making μ go to zero and noting that the second integral goes to zero with μ we find that

$$\frac{\partial F(z_1, z'_1)}{\partial z'_1} = \int_{z_0}^{z_1} \frac{\partial f(z, z'_1)}{\partial z'_1}\,dz.$$

---

[7] Lebesgue, *Annales de l'École Normale*, vol. 27, pp. 363-387.

THEOREM 8. *If $f(z, z')$ is Y-regular then $F(z_1, z'_1)$ is Y-regular.*

We have already seen that

$$\frac{\partial F}{\partial z'_1} = \int_{z_0}^{z_1} \frac{\partial f(z, z'_1)}{\partial z'_1}\, dz$$

while

$$\partial F / \partial z_1 = f(z_1, z'_1).$$

Thus the primary conditions (5.2 a) for Y-regularity are satisfied.

For the secondary conditions we have by (11.5)

$$\frac{\Phi(\lambda, \mu)}{\lambda dz} = \frac{1}{\lambda dz} \int_{z_1}^{z_1+\lambda dz} [f(z, z'_1 + \mu dz') - f(z, z')]\, dz.$$

Here, if μ goes to zero first the integrand becomes identically zero so that

$$\underset{\lambda, \mu \to 0}{\text{limit}} \frac{\Phi(\lambda, \mu)}{\lambda} = 0.$$

If λ goes to zero first one has

$$\underset{\lambda, \mu \to 0}{\text{limit}} \frac{\Phi(\lambda, \mu)}{\lambda} = \underset{\mu \to 0}{\text{limit}}\, f(z_1, z'_1 + \mu dz') - f(z_1, z'_1) = 0.$$

Further

$$\frac{\Phi(\lambda, \mu)}{\mu dz'} = \frac{1}{\mu dz'} \int_{z_0}^{z_1} [f(z, z'_1 + \mu dz') - f(z, z'_1)]\, dz;$$

when λ goes to zero first here, the integral vanishes while if μ goes to zero first

$$\underset{\lambda, \mu \to 0}{\text{limit}} \frac{\Phi(\lambda, \mu)}{\mu dz'} = \underset{\lambda \to 0}{\text{limit}} \int_{z_1}^{z_1+\lambda dz} \frac{\partial f(z_1, z'_1)}{\partial z'_1}\, dz = 0.$$

Thus in all cases the secondary conditions (5.2 b) are satisfied.

THEOREM 9. *If $f(z, z')$ and $g(z, z')$ are each Y-regular their product is Y-regular.*

If this theorem were not true the utility of the definition would be in question. Its proof will be left to the reader with the suggestion that $\phi(f)$ and $\phi(g)$ be constructed. Transpose $f(P)$ and $g(P)$, $P \equiv (z_1 + \lambda dz,\ z'_1 + \mu dz')$ and form the product. From this $\phi(fg)$ may be constructed.

III. TAYLOR SERIES EXPANSION OF A POLYGENIC FUNCTION.

12. Taylor series for $f(z, z')$. Let $f(z, z')$ be R-regular at every point of a neighborhood of $z = z_1$ and of a neighborhood of $z' = z'_1$. Let C and C' be the bounding curves, without double points, of the respective neighborhoods and so chosen that C does not contain the point z'_1 nor C' the point z_1 in their bounded areas. Then it is well known that the Cauchy second law generalizes into [8]

$$f(z, z') = -\frac{1}{4\pi^2}\int_C dt \int_{C'} \frac{f(t, t')\,dt'}{(t-z)(t'-z')} \tag{12.1}$$

where as usual t and t' are points on C and C' respectively. From this it will be evident that

$$\frac{\partial^{m+n}(f(z, z')}{\partial z^m \partial z'^n} = -\frac{m!\,n!}{4\pi^2}\int_C dt \int_{C'} \frac{f(t, t')\,dt'}{(t-z)^{m+1}(t'-z')^{n+1}}. \tag{12.2}$$

We now restrict the contours C and C' to be circles, with centers at z_1 and z'_1 and radii R and R' respectively, and so chosen that neither circle contains the center of the other in its interior or on its boundary. Further

$$\frac{1}{(t-z)(t'-z')} = \frac{1}{t-z_1-(z-z_1)} \cdot \frac{1}{(t'-z'_1)-(z'-z'_1)}$$
$$= \frac{1}{(t-z_1)(t'-z'_1)} \cdot \frac{1}{1-\dfrac{z-z_1}{t-z_1}} \cdot \frac{1}{1-\dfrac{z'-z'_1}{t'-z'_1}}.$$

may be expanded into a power series by simply performing the indicated division in the last two factors and multiplying the resulting series. One finds that

$$\frac{1}{(t-z')(t'-z')} = \frac{1}{(t-z_1)(t'-z'_1)} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \left(\frac{z-z_1}{t-z_1}\right)^m \left(\frac{z'-z'_1}{t'-z'_1}\right)^n. \tag{12.3}$$

If now two other concentric circles radii r and r' are so chosen that

$$|z-z_1| < r < R \quad \text{and} \quad |z'-z'_1| < r' < R'$$

---

[8] Goursat, *Cours d'Analyse*, vol. II, p. 272.

then the series (12. 3) will be uniformly convergent, for

$$\left| \left(\frac{z-z_1}{t-z_1}\right)^m \left(\frac{z'-z'_1}{t'-z'_1}\right)^n \right| \leq \frac{1}{RR'} \left(\frac{r}{R}\right)^m \left(\frac{r'}{R'}\right)^n.$$

From (12. 1) and (12. 3) we find that

$$f(z,z') = -\frac{1}{4\pi^2}\int_C dt \int_{C'} \frac{f(t,t')\,dt'}{(t-z)(t'-z')}$$

$$= -\frac{1}{4\pi^2}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}(z-z_1)^m(z'-z'_1)^n \int_C dt \int_{C'} \frac{f(t,t')\,dt'}{(t-z_1)^{m+1}(t'-z'_1)^{n+1}}$$

and applying (12. 2) to the integral we find that

$$(12.4)\quad f(z,z') = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{1}{m!\,n!}\cdot\frac{\partial^{m+n} f(z_1,z'_1)}{\partial z_1{}^m \partial z'_1{}^n}(z-z_1)^m(z'-z'_1)^n$$

where, by definition, zero factorial is 1 and the zero derivative of the function is the function itself.

If the maximum value of $|f(z,z')|$ on the circles radii r and r' is M then

$$(12.5)\quad \left|\frac{1}{m!\,n!}\;\frac{\partial^{m+n} f(z,z')}{\partial z_1{}^m \partial z'_1{}^n}\right| < \frac{1}{4\pi^2}\;\frac{M}{R^{m+1}R'^{n+1}}\,2\pi R\cdot 2\pi R'. = \frac{M}{R^m R'^n}$$

which is obtained by applying (12. 2) to the integral form of the coefficient. Thus our series (12. 4) is convergent.

13. The Taylor expansion of the polygenic function. We have seen that when $f(z,z')$ is R-regular $f(z,z')$ is given by equation (4. 2). Also when z' is replaced by $\bar{z}$ this becomes

$$f(z) = f(z_1) + \frac{\partial f}{\partial z_1}(z-z_1) + \frac{\partial f}{\partial \bar{z}_1}(\bar{z}-\bar{z}_1) + \eta_1.$$

$\partial f/\partial z_1$ and $\partial f/\partial \bar{z}_1$ are, respectively, $\partial f(z_1,z'_1)/\partial z_1$ and $\partial f(z_1,z'_1)/\partial z'_1$ wherein z'_1 has been replaced by $\bar{z}_1$. It has also been shown that $\partial f/\partial z_1 = \partial f/\partial \beta$, $\partial f/\partial \bar{z}_1 = \partial f/\partial \alpha$ at $z = z_1$ are the areal and mean derivatives respectively. Since all derivatives of $f(z,z')$ are regular it follows that all derivatives of $f(z)$ will be areal and mean derivatives. Hence

$$\partial^{m+n} f/\partial z_1{}^m \partial \bar{z}_1{}^n = \partial^{m+n} f/\partial \beta^m \partial \alpha^n.$$

A substitution of this result into (12. 4) makes

$$(13.1) \qquad f(z) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{1}{m!\,n!} \frac{\partial^{m+n} f(z_1)}{\partial\beta^m \partial\alpha^n} (z - z_1)^m (\bar{z} - \bar{z}_1)^n$$

which is the required expansion of $f(z)$ into a Taylor Series.

(a) The Taylor expansion for $f(z, z')$ includes the case when the domain D' is identified with the conjugate domain $\bar{D}$ of D, making $z'_1 = \bar{z}_1$ the center of the circle C'_1.

(b) When z' is replaced by $\bar{z}$ (12. 1) is no longer valid, i. e. the polygenic function $f(z)$ cannot be expressed in any such form.

(c) The derivative $\partial f(z_1, z'_1)/\partial z'_1$ can be obtained from the function itself and from the double integral in (12. 1), but $\partial f/\partial \bar{z}_1$ can be obtained only from $\partial f(z_1, z'_1)/\partial z'_1$ by putting $z'_1 = \bar{z}_1$. A simliar situation arises for $\partial f/\partial z_1$, of course.

(d) We obtain the right member of (12. 5) from the right member of (12. 2) while the left member may be obtained from the function $f(z_1, z'_1)$ itself.

(e) The point z'_1 is by our geometric set up the very same point as $\bar{z}_1$. Thus the left member of (12. 5) will retain the same value when z'_1 is replaced by $\bar{z}_1$, and therefore the inequality persists. But in this case the left member of (12. 5) becomes a combination of areal and mean derivatives as indicated; The convergence of the series is thus established.

(f) This expansion of a polygenic function into a Taylor series is at least a partial solution of this classical problem. The method used should be of some future value.

THE UNIVERSITY OF MICHIGAN.

CONJUGAL QUADRICS AND THE QUADRIC OF MOUTARD.*

By M. L. MacQueen.

1. Introduction. Among the three-parameter family of quadrics having contact of the second order with a surface at a point, Grove [1; p. 231] has introduced a two-parameter family of quadrics, called *conjugal quadrics,* which are associated with a given conjugate net on the surface. They are found to include certain pencils of quadrics which are of interest, namely, the quadrics of Darboux, the quadrics which have third-order contact [2; p. 421] with both curves of a conjugate net at a point, and the quadrics of Davis [3; p. 12].

The conjugal quadrics which are associated with the curves of a conjugate net at a point P_x of a surface and the quadric of Moutard in the direction λ of a tangent of a curve of the net intersect in the asymptotic tangents through the point P_x and in a conic which lies in a plane through P_x. It is the purpose of this paper to study the envelope of this plane when λ varies.

In **2**, we are concerned with the calculation of the equations of the envelope, which is found to be a cone of the sixth class with its vertex at the point P_x. A study of the envelope and its polars, in **3**, enables us to obtain characterizations of certain conjugal quadrics which appear to be of some interest. Among other things, we present several geometric definitions of the general canonical line of the first kind. The last four sections, which are devoted to a brief study of special cases, contain geometric characterizations of unique conjugal quadrics and new interpretations of particular canonical lines.

In a recent paper, Chang [4; p. 926] studied the envelope of the plane containing the residual conic of intersection of the quadrics of Darboux and the quadric of Moutard for a direction λ at a point P_x when λ varies. Consequently, it is of interest to obtain some of his results as a special case of the problem considered here.

2. The envelope. Let the differential equations of a non-ruled surface S in ordinary projective space be written in the Fubini canonical form [5; p. 69]

$$x_{uu} = px + \theta_u x_u + \beta x_v, \qquad x_{vv} = qx + \gamma x_u + \theta_v x_v \tag{2.1}$$
$$(\theta = \log \beta\gamma).$$

* Received March 20, 1945.

We select an ordinary point P_x of the surface S as one vertex of the usual local tetrahedron of reference x, x_u, x_v, x_{uv}. A conjugate net N_λ on the surface S can be represented by a curvilinear differential equation of the form

$$dv^2 - \lambda^2 du^2 = 0 \qquad (\lambda \neq 0), \tag{2.2}$$

where λ is a function of u, v. The two curves of the net N_λ that pass through the point P_x may be denoted by C_λ and $C_{-\lambda}$ according as the direction dv/du has the value λ or $-\lambda$.

With the introduction of nonhomogeneous projective coördinates x, y, z defined in the customary way, the equation of the conjugal quadrics associated with the net N_λ at the point P_x can be written in the form

$$xy - z + h(\gamma\lambda^2 xz + \beta yz/\lambda^2) + lz^2 = 0, \tag{2.3}$$

in which h, l are arbitrary parameters.

Among the conjugal quadrics, perhaps the most important are those for which h is a constant. Such quadrics include the quadrics of Darboux with $h = 0$, the quadrics of Davis with $h = 1$, the quadrics with $h = -\frac{1}{3}$ which have third-order contact with both curves of the net N_λ, and the quadrics corresponding to various constant values of h for which Grove has given geometric characterizations [1, p. 233].

We shall suppose from now on that h is a constant which is independent of u and v.

The quadric of Moutard for a direction λ at the point P_x of the surface has the equation

$$\begin{aligned} &36\lambda^3(xy - z) + 12\lambda^2(\gamma\lambda^4 - 2\beta)xz - 12\lambda(2\gamma\lambda^3 - \beta)yz \\ &\quad + [3\beta\phi\lambda - 12\beta\psi\lambda^2 - 18\theta_{uv}\lambda^3 - 12\gamma\phi\lambda^4 + 3\gamma\psi\lambda^5 - 4(\beta + \gamma\lambda^3)^2]z^2 = 0, \end{aligned} \tag{2.4}$$

where the functions ϕ and ψ are defined by

$$\phi = (\log \beta\gamma^2)_u, \qquad \psi = (\log \beta^2\gamma)_v.$$

The two quadrics (2.3) and (2.4) intersect in a curve consisting of the asymptotic tangents at the point P_x of the surface, and in a conic which lies in the plane π_λ whose equation is

$$\begin{aligned} &12\lambda^2(2\beta - \gamma\lambda^3 + 3h\gamma\lambda^3)x + 12\lambda(3h\beta - \beta + 2\gamma\lambda^3)y \\ &\quad + [4\beta^2 - 3\beta\phi\lambda + 12\beta\psi\lambda^2 + (8\beta\gamma + 36L\beta\gamma)\lambda^3 + 12\gamma\phi\lambda^4 \\ &\quad - 3\gamma\psi\lambda^5 + 4\gamma^2\lambda^6]z = 0, \end{aligned} \tag{2.5}$$

in which we have placed

$$= L\beta\gamma - \tfrac{1}{2}\theta_{uv}. \tag{2.6}$$

The homogeneous plane coördinates of the plane π_λ are given by the equations

$$
\begin{aligned}
&\rho u_1 = 0,\\
&\rho u_2 = 24\beta\lambda^2 - 12(1-3h)\gamma\lambda^5,\\
&\rho u_3 = -12(1-3h)\beta\lambda + 24\gamma\lambda^4,\\
&\rho u_4 = 4\beta^2 - 3\beta\phi\lambda + 12\beta\psi\lambda^2 + (8\beta\gamma + 36L\beta\gamma)\lambda^3 + 12\gamma\phi\lambda^4\\
&\qquad - 3\gamma\psi\lambda^5 + 4\gamma^2\lambda^6,
\end{aligned}
\tag{2.7}
$$

where ρ is a proportionality factor not zero.

The envelope of the plane π_λ when λ varies may be found by homogeneous elimination of λ from equations (2.7). For convenience, we follow Chang [4; p. 927] and obtain, after some calculation based on equations (2.7), an equivalent system of equations, namely,

$$
\begin{aligned}
&\pi\lambda^3 + \omega\lambda^2 + t\lambda + s = 0,\\
&\sigma\lambda^3 + \tau\lambda^2 + \omega\lambda + p = 0,
\end{aligned}
\tag{2.8}
$$

whose coefficients are given by the following formulas which preserve Chang's notation:

$$
\begin{aligned}
\pi &= -\frac{2(1+3h)(1-h)}{1-3h}\gamma u_2 + (1-6h)\psi u_3,\\
\omega &= \tfrac{1}{2}(7+3h)\psi u_2 + \tfrac{1}{2}(7+3h)\phi u_3 - 6(1+3h)(1-h)u_4;\\
t &= (1-6h)\phi u_2 + 6\left[\frac{(1-h)^2 + 2(1-3h)L}{1-3h}\right]\beta u_3,\\
s &= [4(1-h) + 6(1-3h)L]\beta u_2,\\
\sigma &= [4(1-h) + 6(1-3h)L]\gamma u_3,\\
\tau &= (1-6h)\psi u_3 + 6\left[\frac{(1-h)^2 + 2(1-3h)L}{(1-3h)}\right]\gamma u_2;\\
p &= (1-6h)\phi u_2 - \frac{2(1+3h)(1-h)}{1-3h}\beta u_3,
\end{aligned}
\tag{2.9}
$$

in which we have assumed $h \neq 1$ and $h \neq \pm\frac{1}{3}$. These excluded cases will be considered in the last three sections.

If we eliminate λ from equations (2.8), we find that *the envelope of the plane π_λ is the cone whose equations in local plane coördinates are $u_1 = 0$ and*

$$
\begin{vmatrix}
\pi & \omega & t & s\\
\sigma & \tau & \omega & p\\
0 & p\pi - \sigma s & \omega p - \tau s & pt - \omega s\\
\omega\sigma - \pi\tau & \sigma t - \omega\pi & \sigma s - \pi p & 0
\end{vmatrix} = 0.
\tag{2.10}
$$

In the first place, if L has the value given by

$$
L = -\frac{2}{3}\left(\frac{1-h}{1-3h}\right),
\tag{2.11}
$$

equations (2.9) show that $t = p$, $\tau = \pi$, and $s = \sigma = 0$, in consequence of which it is not difficult to verify that the determinant appearing in equation (2.10) is identically zero. From this fact we infer that the envelope of the plane π_λ differs from the cone (2.10), in case L is defined by equation (2.11). Consequently we shall assume, for the present, that the parameter L is arbitrary and is not given by (2.11).

3. Properties of the envelope. The third polar of the tangent plane $z = 0$ of the surface at the point P_x with respect to the cone (2.10) is a cone of class three which is found to have the equations $u_1 = 0$ and

$$(3.1) \qquad \begin{aligned} 6u_2u_3[u_4 &- \frac{(2-4h-3h^2)}{4(1+3h)(1-3h)(1-h)}\phi u_3 \\ &- \frac{(2-4h-3h^2)}{4(1+3h)(1-3h)(1-h)}\psi u_2] \\ &- \frac{1}{(1-3h)^2}(\beta u_3^3 + \gamma u_2^3) = 0. \end{aligned}$$

It will be observed that this equation is independent of L, so that *all of the cones* (2.10) *with different values of the parameter L have in common the same third polar with respect to the tangent plane $z = 0$.*

The cusp-axis of the cone (3.1) is the canonical line $l_1(k)$ for which

$$(3.2) \qquad k = -\frac{(2-4h-3h^2}{4(1+3h)(1-3h)(1-h)}.$$

Thus we have a construction which yields the general canonical line of the first kind and which may be described in the following theorem:

If the parameter L is not defined by (2.11), *corresponding to each conjugal quadric* (2.3), *except those with $h = 1$ and $h = \pm\frac{1}{3}$, there is a canonical line $l_1(k)$ for which k is defined by the formula* (3.2), *and this line is the cusp-axis of the cone* (3.1).

In particular, it is perhaps worthy of remark that the conjugal quadrics with $h = 0$, $h = 1/6$, and $h = 2/3$ all correspond in this manner to the first directrix of Wilczynski.

The second polar of the tangent plane $z = 0$ with respect to the cone (2.10) is a cone of the fourth class with the equations $u_1 = 0$ and

$$(3.3) \qquad \begin{aligned} 6u_2u_3\omega^2 - 6 &\left[\frac{(1-6h)}{(1-3h)}(\psi u_2 + \phi u_3)u_2u_3\right. \\ &\left. - \frac{2(1+3h)(1-h)}{(1-3h)^2}(\beta u_3^3 + \gamma u_2^3)\right]\omega + [\ldots]u_2u_3 = 0, \end{aligned}$$

where $[\ldots]$ is independent of ω but does contain the parameter L.

The equation of any plane through the v-tangent $x=z=0$, distinct from the tangent plane, can be written in the form

$$x+az=0, \tag{3.4}$$

in which a is a parameter. If the homogeneous plane coördinates of this plane are substituted in equation (3.3), we find that the plane (3.4) is tangent to the cone (3.3) if, and only if,

$$a=\frac{7+3h}{12(1+3h)(1-h)}\psi. \tag{3.5}$$

Similarly, a plane through the u-tangent $y=z=0$ has the equation

$$y+bz=0,$$

in which b is a parameter, and is tangent to the cone (3.3) in case

$$b=\frac{7+3h}{12(1+3h)(1-h)}\phi. \tag{3.6}$$

Thus we prove the following theorem:

Through each asymptotic tangent at the point P_x of the surface S there is a plane which is tangent to all of the cones (3.3) with variable L. These two planes intersect in the canonical line $l_1(k)$ for which

$$k=-\frac{7+3h}{12(1+3h)(1-h)}. \tag{3.7}$$

The equation of any plane, except the tangent plane $z=0$, through the tangent line in the direction λ at the point P_x of the surface is

$$y-\lambda x=\mu z, \tag{3.8}$$

where μ is a parameter. If the plane coördinates of this plane are substituted in equation (3.3), we find

$$A\mu^2+B\mu+C=0, \tag{3.9}$$

in which

$$\begin{aligned} A&=12(1+3h)(1-h)\lambda,\\ B&=\frac{3(2-4k-3h^2)}{1-3h}(\phi-\psi\lambda)\lambda-\frac{2(1+3h)(1-h)}{(1-3h)^2}(\beta-\gamma\lambda^3), \end{aligned} \tag{3.10}$$

the definition of C not being essential to our work. Hence, through the tangent line in the direction λ there are two planes p_1, p_2, besides the tangent plane of the surface, which are tangent to the cone (3.3). The harmonic conjugate of the tangent plane $z = 0$ with respect to the two planes p_1, p_2 is the plane p^* whose coördinates are given by

$$(3.11) \qquad \rho u_1 = 0, \qquad \rho u_2 = -A\lambda, \qquad \rho u_3 = A, \qquad \rho u_4 = B.$$

The plane p^* is tangent to the cone (3.1) if, and only if,

$$(3.12) \qquad 2h(1+3h)(1-h)(\beta-\gamma\lambda^3) \\ -3h(1-3h)(2-4h-3h^2)(\phi-\psi\lambda)\lambda = 0.$$

This condition is obviously satisfied if $h = 0$, that is, if the conjugal quadric (2.3) is a quadric of Darboux. If $h \neq 0$, equation (3.12) represents the triple of directions D^*_k which are conjugate to the D_k directions introduced, respectively, by Wilkins [6; p. 177] and Bell [7; p. 787]. Moreover, condition (3.12) is satisfied if $\phi - \psi\lambda = 0$, $\beta - \gamma\lambda^3 = 0$, so that the curve C_λ is a curve of Segre which is tangent to the first canonical tangent at P_x. These results may be summarized in the following statement:

There are two planes p_1, p_2, besides the tangent plane of the surface S, through the tangent line in the direction λ at the point P_x which are tangent to the cone (3.3). The plane p^* which is the harmonic conjugate of the tangent plane $z = 0$ with respect to the planes p_1 and p_2 envelopes the cone (3.1) in case any one of the following conditions holds:

(i) *the conjugal quadric (2.3) is a quadric of Darboux;*

(ii) *the direction of the curve C_λ is a direction D^*_k;*

(iii) *the curve C_λ is a curve of Segre which is tangent to the first canonical tangent at P_x.*

Moreover, it is not difficult to verify the truth of the following theorem:

At a point P_x of the surface S, let p_1, p_2 denote the two planes through the tangent line in the direction λ at the point P_x which are tangent to the cone (3.3). Let p^* denote the plane which is the harmonic conjugate of the tangent plane $z = 0$ with respect to the planes p_1, p_2. Then, the three planes

p^*_i $(i = 1, 2, 3)$ *corresponding to the Segre tangents at* P_x *are concurrent in the canonical line* $l_1(k)$ *for which* k *is given by the formula* (3.2).

4. The case when $L = -(2/3)[(1-h)/(1-3h)]$. In this section we assume that L is defined by equation (2.11). Equation (2.3) becomes, in consequence of (2.6) and (2.11),

$$(4.1)\quad xy - z + h(\gamma\lambda^2 xz + \beta yz/\lambda^2) - \left[\frac{2}{3}\left(\frac{1-h}{1-3h}\right)\beta\gamma + \tfrac{1}{2}\theta_{uv}\right]z^2 = 0,$$

where h is a constant which is independent of u and v. Thus, *the conjugal quadrics* (4.1) *are characterized by the property that the envelope of the plane* (2.5) *is different from the cone* (2.10)

It may be remarked that the quadric (4.1) with $h = 0$ is the quadric of Darboux which Chang introduced [4; p. 928].

We now proceed to find the envelope of the plane π'_λ for the quadrics represented by equation (4.1). For this purpose we replace equations (2.8) by the equations

$$(4.2)\quad \begin{aligned} &[(1-3h)\gamma\omega u_3 - 2\gamma\pi u_2]\lambda^3 + (1-3h)\gamma p u_3\lambda^2 \\ &\qquad\qquad + 2\beta\pi u_3\lambda + (1-3h)\beta\pi u_2 = 0, \\ &\pi\lambda^2 + \omega\lambda + p = 0, \end{aligned}$$

in which ω, π, and p are given by

$$(4.3)\quad \begin{aligned} \omega &= \tfrac{1}{2}(7+3h)\psi u_2 + \tfrac{1}{2}(7+3h)\phi u_3 - 6(1+3h)(1-h)u_4, \\ \pi &= -\frac{2(1+3h)(1-h)}{(1-3h)}\gamma u_2 + (1-6h)\psi u_3, \\ p &= (1-6h)\phi u_2 - \frac{2(1+3h)(1-h)}{(1-3h)}\beta u_3. \end{aligned}$$

Elimination of λ from equations (4.2) shows that *the envelope of the plane* π'_λ *is a cone of the sixth class whose equations in plane coördinates are* $u_1 = 0$ *and*

$$(4.4)\quad \begin{vmatrix} (1-3h)\gamma p u_3 & 2\gamma p u_2 & (1-3h)\beta\pi u_2 & (1-3h)\beta\omega u_2 - 2\beta p u_3 \\ (1-3h)\gamma\omega u_3 - 2\gamma\pi u_2 & (1-3h)\gamma p u_3 & 2\beta\pi u_3 & (1-3h)\beta\pi u_2 \\ \pi & \omega & p & 0 \\ 0 & \pi & \omega & p \end{vmatrix} = 0$$

The second polar of the tangent plane $z = 0$ with respect to the cone (4.4) has the equations $u_1 = 0$ and

$$(4.5) \qquad (1-3h)\pi\omega u_2^2 + (1-3h)p\omega u_3^2 - (1-3h)^2\omega^2 u_2 u_3 + 2h(3h-2)\pi p u_2 u_3 = 0.$$

If $h=0$, the second polar (4.5) decomposes into a cone of the third class and a pencil of planes whose axis is the canonical line $l_1(k)$ for which $k=-7/12$, as Chang has shown. Moreover, if $h=2/3$, the second polar also decomposes into a cone of the third class and a pencil of planes whose axis is the canonical line $l_1(k)$ for which $k=-3/4$. We summarize these results in the following statement:

Among the quadrics (4.1), *there are two quadrics, namely, those with* $h=0$ *and* $h=2/3$, *which are characterized by the property that the corresponding cone* (4.5) *decomposes into a cone of the third class and a pencil of planes. The axes of these two pencils of planes are the canonical lines* $l_1(k)$ *for which* k *has the respective values* $k=-7/12$ *and* $k=-3/4$.

If we exclude the two values of h for which the cone (4.5) is composite, it is easy to show that *the tangent planes of the cone* (4.5) *which pass through the asymptotic tangents at the point* P_x *intersect in the canonical line* $l_1(k)$ *for which* k *is given by the formula* (3.7).

The third polar of the tangent plane $z=0$ with respect to the cone (4.4) is the cone (3.1). Finally, the fourth polar of the tangent plane $z=0$ with respect to the cone (4.4) consists of two pencils of planes each having the asymptotic tangents at P_x for an axis.

5. The case when $h=-1/3$. It has been remarked that equation (2.3) with $h=-1/3$ defines a pencil of quadrics each of which has third order contact with the curves of the net N_λ at P_x.

In this case the homogeneous plane coördinates of the plane (2.5) are

$$(5.1) \qquad \begin{aligned} \rho u_1 &= 0, \\ \rho u_2 &= 24\lambda^2(\beta-\gamma\lambda^3), \\ \rho u_3 &= -24\lambda(\beta-\gamma\lambda^3), \\ \rho u_4 &= 4\beta^2 - 3\beta\phi\lambda + 12\beta\psi\lambda^2 + (8\beta\gamma + 36L\beta\gamma)\lambda^3 + 12\gamma\phi\lambda^4 \\ &\quad - 3\gamma\psi\lambda^5 + 4\gamma^2\lambda^6. \end{aligned}$$

Homogeneous elimination of λ from these equations yields *a cone of the sixth class whose equations are*

$$(5.2) \qquad \begin{aligned} &4(\beta u_3^3 + \gamma u_2^3)(\beta u_3^3 + \gamma u_2^3 + \tfrac{3}{4}\phi u_2 u_3^2 + \tfrac{3}{4}\psi u_2^2 u_3 - 6u_2u_3u_4) \\ &\quad + 9\beta\psi u_2^2 u_3^4 + 9\gamma\phi u_2^4 u_3^2 - (16\beta\gamma + 36L\beta\gamma)u_2^3u_3^3 = 0, \quad u_1 = 0. \end{aligned}$$

The following statement may be easily verified.

The first polar of planes through the asymptotic u-tangent (v-tangent) at the point P_x with respect to the cone (5. 2) contains a plane $p_1(p_2)$ which passes through the v-tangent (u-tangent) at P_x. The intersection of the planes p_1 and p_2 is the canonical line $l_1(k)$ for which $k = -1/8$.

Moreover, the first polar of planes through the u-tangent at P_x with respect to the cone (5. 2) contains a plane through each of the three tangents of Segre at P_x. The equations of these planes π_i $(i = 1, 2, 3)$ are

$$(5.3) \quad \begin{aligned} &24(4\beta - \gamma\lambda^3)(\lambda x - y) + [24\beta^2/\lambda - 15\beta\phi \\ &\quad + 48\beta\psi\lambda + 3(8\beta\gamma + 36L\beta\gamma)\lambda^2 + 24\gamma\phi\lambda^3 - 3\gamma\psi\lambda^4]z = 0, \end{aligned}$$

where

$$\lambda = \omega^i(\beta/\gamma)^{1/3}; \quad i = 1, 2, 3; \quad \omega^3 = 1, \ \omega \neq 1.$$

The polar line of the tangent plane $z = 0$ with respect to the trihedron formed by the planes π_i is the line l_1 which joins P_x to the point whose local point coördinates are

$$(5.4) \qquad (0, -\tfrac{5}{8}\psi, \tfrac{1}{8}\phi, 1).$$

In a similar manner, we find that the first polar of planes through the v-tangent with respect to the cone (5. 2) contains a plane through each of the tangents of Segre at P_x. These three planes π^*_i $(i = 1, 2, 3)$ are given by the equations

$$(5.5) \quad \begin{aligned} &24(4\gamma\lambda^3 - \beta)(\lambda x - y) + [3\beta\phi - 24\beta\psi\lambda - 3(8\beta\gamma \\ &\quad + 36L\beta\gamma)\lambda^2 - 48\gamma\phi\lambda^3 + 15\gamma\psi\lambda^4 - 24\gamma^2\lambda^5]z = 0, \end{aligned}$$

where

$$\lambda = \omega^i(\beta/\gamma)^{1/3}; \quad i = 1, 2, 3; \quad \omega^3 = 1, \ \omega \neq 1.$$

The polar line of the tangent plane $z = 0$ with respect to the trihedron formed by the planes π^*_i is the line l^*_1 joining P_x to the point

$$(5.6) \qquad (0, \tfrac{1}{8}\psi, -\tfrac{5}{8}\phi, 1).$$

The plane determined by the lines l_1 and l^*_1 thus defined intersects the canonical plane in the first edge of Green.

We present a geometric characterization of a unique conjugal quadric (2. 3) with $h = -1/3$ which is contained in the following theorem:

The third polar of planes through the u-tangent (v-tangent) with respect to the cone (5.2) contains a plane which passes through the v-tangent (u-tangent) if, and only if, $L = -2/9$. In this case the quadric has the equation

$$(5.7) \qquad xy - z - \tfrac{1}{3}(\gamma\lambda^2 xz + \beta yz/\lambda^2) - (\tfrac{2}{9}\beta\gamma + \tfrac{1}{2}\theta_{uv})z^2 = 0.$$

Finally, direct calculation, which will be omitted, yields the following result:

At a point P_x of a surface, the fourth polar of planes through the u-tangent (v-tangent) with respect to the cone (5.2) is independent of L and contains a plane which passes through the v-tangent (u-tangent). The line of intersection of these two planes is the first directrix of Wilczynski.

6. The case when $h = 1/3$. The quadric (2.3) with $h = 1/3$ is related to the associate conjugate net of the net N_λ in the same way that the quadric with $h = -1/3$ is related to the net N_λ.

The plane coördinates of the plane (2.5) are

$$(6.1) \qquad \begin{aligned} &\rho u_1 = 0, \qquad \rho u_2 = 24\beta\lambda^2, \qquad \rho u_3 = 24\gamma\lambda^4, \\ &\rho u_4 = 4\beta^2 - 3\beta\phi\lambda + 12\beta\psi\lambda^2 + (8\beta\gamma + 36L\beta\gamma)\lambda^3 + 12\gamma\phi\lambda^4 \\ &\qquad - 3\gamma\psi\lambda^5 + 4\gamma^2\lambda^6. \end{aligned}$$

The envelope of this plane when λ varies is found to be *a cone of the sixth class whose equations are*

$$(6.2) \qquad \begin{aligned} &16\beta\gamma[\beta u_3^3 + \gamma u_2^3 - 3(2u_4 - \phi u_3 - \psi u_2)u_2u_3]^2 \\ &\quad - [3(\beta\psi u_3^3 + \gamma\phi u_2^2) - (8\beta\gamma + 36L\beta\gamma)u_2u_3]^2 u_2u_3 = 0, \quad u_1 = 0. \end{aligned}$$

The first polar of the tangent plane $z = 0$ with respect to this cone decomposes into a cone of the third class

$$(6.3) \qquad \beta u_3^3 + \gamma u_2^3 - 3(2u_4 - \phi u_3 - \psi u_2)u_2u_3 = 0, \quad u_1 = 0,$$

and two pencils of planes each having an asymptotic tangent as axis. It will be observed that *the cusp-axis of the cone (6.3) is the first directrix of Wilczynski.*

There are two planes p_1, p_2 through the tangent line in the direction λ

at the point P_x which are tangent to the cone (6.2). The harmonic conjugate of the tangent plane $z=0$ with respect to the planes p_1, p_2 is the plane p^* whose equation is

$$(6.4) \qquad 6\lambda(x_3-\lambda x_2)+[3(\phi-\psi\lambda)\lambda-(\beta-\gamma\lambda^3)]x_4=0.$$

The planes p^*_i $(i=1,2,3)$ corresponding to the three tangents of Segre are concurrent in the first directrix of Wilczynski.

7. The case when $h=1$. The quadrics of Davis for the net N_λ are the quadrics (2.3) with $h=1$. In this case the plane coördinates of the plane (2.5) are

$$(7.1) \qquad \begin{aligned} &\rho u_1=0, \qquad \rho u_2=24\lambda^2(\beta+\gamma\lambda^3), \qquad \rho u_3=24\lambda(\beta+\gamma\lambda^3),\\ &\rho u_4=4\beta^2-3\beta\phi\lambda+12\beta\psi\lambda^2+(8\beta\gamma+36L\beta\gamma)\lambda^3+12\gamma\phi\lambda^4\\ &\qquad -3\gamma\psi\lambda^5+4\gamma^2\lambda^6. \end{aligned}$$

The envelope of this plane is *a cone of the sixth class with the equations*

$$(7.2) \qquad \begin{aligned} &4(\beta u_3^3+\gamma u_2^3)(\beta u_3^3+\gamma u_2^3-\tfrac{3}{4}\phi u_2u_3^2-\tfrac{3}{4}\psi u_2^2u_3-6u_2u_3u_4)\\ &\quad +15\beta\psi u_2^2u_3^4+15\gamma\phi u_2^4u_3^2+36L\beta\gamma u_2^3u_3^3=0, \quad u_1=0. \end{aligned}$$

It is not difficult to show that the first polar of planes through the u-tangent at P_x with respect to the cone (7.2) contains a plane through each of the tangents of Segre at P_x. The polar line of the tangent plane $z=0$ with respect to the trihedron formed by these three planes is the line l_1 which joins P_x to the point

$$(7.3) \qquad (0, \ -{}^{17}\!/_{24}\psi, \ \ {}^{13}\!/_{24}\phi, \ \ 1).$$

Similarly, the first polar of planes through the v-tangent with respect to the cone (7.2) contains a plane through each of the Segre tangents. The polar line of the tangent plane $z=0$ with respect to this trihedron is the line l^*_1 which joins P_x to the point

$$(7.4) \qquad (0, \ \ {}^{13}\!/_{24}\psi, \ -{}^{17}\!/_{24}\phi, \ \ 1).$$

The plane determined by the lines l_1, l^*_1 intersects the canonical plane in the canonical line $l_1(k)$ for which $k=-1/12$.

Finally, direct calculation yields the following result:

The third polar of planes through the u-tangent (v-tangent) with

respect to the cone (7.2) *contains a plane which passes through the v-tangent* (*u-tangent*) *if, and only if,* $L = -2/9$. *Consequently, we have a geometric characterization of the unique quadric represented by the equation*

$$xy - z + \gamma\lambda^2 xz + \beta yz/\lambda^2 - (\tfrac{2}{9}\beta\gamma + \tfrac{1}{2}\theta_{uv})z^2 = 0. \tag{7.5}$$

SOUTHWESTERN COLLEGE,
MEMPHIS, TENN.

REFERENCES.

1. V. G. Grove, "The transformation of Čech," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 231-234.

2. M. L. MacQueen, "Pencils of quadrics associated with a conjugate net," *Journal of the Tennessee Academy of Science*, vol. 15 (1940), pp. 420-423.

3. W. M. Davis, *Contributions to the Theory of Conjugate Nets*, Chicago doctoral dissertation, 1932.

4. S. C. Chang, "On the quadrics associated with a point of a surface," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 926-930.

5. E. P. Lane, *Projective Differential Geometry of Curves and Surfaces*, Chicago, 1932.

6. J. E. Wilkins, "The first canonical pencil," *Duke Mathematical Journal*, vol. 10 (1943), pp. 173-178.

7. P. O. Bell, "The first canonical pencil,' *Duke Mathematical Journal*, vol. 5 (1939), pp. 784-788.

THE INFINITIES IN THE NON-LOCAL EXISTENCE PROBLEM OF ORDINARY DIFFERENTIAL EQUATIONS.*

By AUREL WINTNER.

The standard general existence theorem for the problem of initial values assigned for systems of ordinary differential equations is strictly local in nature. There seem to be just two standard types which are particular enough to allow a control of the solutions in the large. The first of these classical cases is represented by systems of linear differential equations, the second by those systems the solutions of which depend on inversions of quadratures.

This latter type is exemplified, in the complex domain, by the differential equation $dp/dz = (p^3 + ap + b)^{\frac{1}{2}}$ of the elliptic function $p(z)$ and, in the real domain, by Liouville's separable conservative systems with n degrees of freedom. In fact, if x denotes any of the n separating coordinates, the energy relation of each degree of freedom is a differential equation of the form $dx/dt = f(x)$. As illustrated by the example $f(x) = x^{1/3}$, where $-\infty < x < \infty$, the mere continuity of $f(x)$ is insufficient for the uniqueness of the solution $x(t)$ belonging to a given initial value $x(0)$, since $x(t) = \alpha \mid t \mid^{3/2} \operatorname{sgn} t$ then is a solution of $dx/dt = f(x)$, $x(0) = 0$, for *two* non-vanishing values of α (and also for $\alpha = 0$).

However, since the zeros of a continuous function form a closed set (and possess, therefore, a complement consisting of a sequence of mutually disjoint open intervals), it is easily realized that, no matter how complicated the real, continuous function $f(x)$, $-\infty < x < \infty$, may be, *a solution $x(t)$ of $dx/dt = f$ cannot cease to exist at a finite $t = t_0$ without tending either to $+\infty$ or to $-\infty$ as $t \to t_0$.* (That such a value t_0 can actually occur, is shown by the movable singularity of the solutions $x(t) = (t_0 - t)^{-1}$ of the differential equation $dx/dt = x^2$, in which $f(x)$ is regular throughout.) In the elliptic case of the complex field, the italicized remark manifests itself in the fact that the only points z $(\neq \infty)$ at which $w = p(z + \text{const.})$ fails to satisfy its (not rational, though algebraic) differential equation $dw/dz = f(w)$ are infinities proper, since all of its singularities are poles, and so, in particular, neither of the type $(z - z_0)^{\frac{1}{2}}$ nor of the type $\exp(z - z_0)^{-1}$. And all singu-

* Received June 11, 1945; Revised October 12, 1945.

larities of $w(z)$ (for $z \neq \infty$) remain infinities proper in all hyperelliptic cases of $dw/dz = f(w)$. In this regard, it is besides the point that there are no branch points at all, not even infinities proper of the type $(z - z_0)^{-\frac{1}{2}}$, in the unramified case. With regard to branch points of infinite order, it is understood that a logarithmic singularity represents an infinity proper, if it is of the type $\log(z - z_0)$, but does not, if it is of the type $(z - z_0)\log(z - z_0)$ or $1/\log(z - z_0)$.

What concerns the first of the two standard cases mentioned at the very beginning, it was shown in [3] that, by a repeated application of the local existence theorem, it is possible to delimit a class of non-linear systems for which the behavior of all solutions in the large is typified by the classical case of linear systems. In the present note, there will be delimited a class of differential equations for which the other case, that depending on inversions of quadratures, is the prototype. By this is implied that, in the real domain, the property italicized above will hold for all solutions. In the complex domain, this must be replaced by the property that all (finite) singularities of all solutions are infinities proper in the sense described above. The principal device will consist of an adaptation of that application of the local existence theorem of arbitrary ordinary differential equations which was used by Painlevé both in the complex field (cf., e. g., [2], p. 23) and in his discussion of the real singularities of the problem of three bodies (*ibid.*, pp. 584-586).

In the real domain, the theorem to be proved becomes the following fact:

If $f(t, x)$ is a real-valued, continuous function on a strip

$$0 \leqq t \leqq a, \qquad -\infty < x < \infty,$$

and if a real-valued function

$$x = x(t), \qquad 0 \leqq t < t_0,$$

where $t_0 \leqq a$ (and $t_0 \neq \infty$), is a solution of the differential equation

$$x' = f(t, x),$$

then, as t tends from the left to the excluded value t_0,

either there exists a finite limit $x(t_0 - 0)$, in which case the function $x(t)$ can be extended to the closed interval $0 \leqq t \leqq t_0$ (and, if $t_0 \neq a$, to an interval $0 \leqq t \leqq t^*$, where $t_0 < t^* < a$) in such a way as to possess a continuous derivative and to satisfy the differential equation,

or else $|x(t)|$ *tends to* ∞, *which, since* $x(t)$ *is continuous for* $0 \leqq t < t_0$, *means that the limit* $x(t_0 - 0)$ *exists either as* $+\infty$ *or as* $-\infty$.

In particular, $\liminf |x(t)| < \infty$ is impossible unless $\text{lim-sup}\, |x(t)| < \infty$, as $t \to t_0 - 0$. Actually, what the theorem really claims is that $\liminf |x(t)| < \infty$ is impossible unless there exists a (finite) $\lim x(t)$, as $t \to t_0 - 0$. In fact, it is easy to see from well-known general properties of functions which are derivatives (and then, if $t_0 \neq a$, from the local existence theorem of ordinary differential equations), that this particular claim, though just a corollary, happens to be equivalent to the theorem.

Correspondingly, the content of the theorem can be illustrated as follows: No matter what the continuous function $f(t, x)$ (*on a strip*) may be, a solution $x(t)$ cannot behave as

$$\sin (t - t_0)^{-1} \quad \text{or} \quad (t - t_0)^{-1} \sin (t - t_0)^{-1},$$

since both of these functions satisfy $\liminf |x(t)| < \infty$ but possess no $\lim x(t)$, as $t \to t_0 - 0$. In addition, the behavior of

$$(t - t_0) \sin (t - t_0)^{-1} \quad \text{or} \quad (t - t_0)^2 \sin (t - t_0)^{-1}$$

is excluded, since, though both of these functions are continuous (if defined to be 0 at $t = t_0$), the first of them has no derivative, and the second no continuous derivative, at $t = t_0$.

It is easy to realize that generalizations for the case of an order higher than the first are not possible in certain directions. For instance, one of the simplest cases not covered by the theorem is a system of $n = 2$ differential equations of the first order not containing the independent variable, say

$$dx/dt = f(x, y), \qquad dy/dt = g(x, y),$$

where both functions f, g are regular on the whole (x, y)-plane. But the assertions of the theorem may then become false, even if both functions are restricted to be polynomials. This is proved by the example (cf. [2], p. 545)

$$f = y(x^2 + y^2)^2 + \tfrac{1}{2}x(x^2 + y^2), \qquad g = x(x^2 + y^2)^2 - \tfrac{1}{2}y(x^2 - y^2),$$

since, as is seen by differentiations and substitutions,

$$x(t) = (t_0 - t)^{-\frac{1}{2}} \sin (t_0 - t)^{-1}, \qquad y(t) = (t_0 - t)^{-\frac{1}{2}} \cos (t_0 - t)^{-1}$$

is then a (real) solution (for $-\infty < t < t_0$).

Let $x(t)$, where $0 \leqq t < t_0$, be a solution satisfying $\liminf |x(t)| < \infty$ as $t \to t_0 - 0$. This means that there exist an increasing sequence $t_1, t_2, \cdots$ and a constant c satisfying $t_m \to t_0$ and $|x(t_m)| \leqq c$. Corresponding to this c and to an arbitrary positive constant b, choose an $M = M(b, c)$ so large that the inequality $|f(t, x)| \leqq M$ holds at every point of the rectangle

$$0 \leqq t \leqq t, \qquad -b - c \leqq x \leqq b + c.$$

The existence of such an M is assured by the assumption that $f(t, x)$ is given as a continuous function on the strip $0 \leqq t \leqq a$ of the (t, x)-plane.

The rectangle defined by the last formula line contains the rectangle

$$t_m \leqq t \leqq a, \qquad -b \leqq x - x(t_m) \leqq b$$

for every m, since $0 \leqq t_m < t_0 \leqq a$ and $|x(t_m)| \leqq c$. Hence, $|f(t, x)| \leqq M$ holds at every point of the rectangle just mentioned, where m is arbitrary. It follows therefore from the local existence theorem of ordinary differential equations, that $x' = f(t, x)$ has on the closed t-interval

$$t_m \leqq t \leqq t_m + \min(a - t_m, b/M)$$

a solution $x(t)$ attaining the given initial value $x(t_m)$ at $t = t_m$. Accordingly, in order to complete the proof of the theorem, it is sufficient to ascertain that, if m is large enough, the upper end-point, $t = t_m + \min$, of this closed t-interval is beyond or at the point $t = t_0$ according as $t_0 < a$ or $t_0 = a$. But this is implied by the assumption $t_m \to t_0$, since b/M is independent of m.

Appendix.

By an adaptation of the proof given above, it is possible to obtain, in a generalized form, a simple proof of the principal result of [3]. The generalization consists in omitting an assumption of monotony for the function $L(r)$ below. The resulting theorem, which is a complete dual of Osgood's criterion for the uniqueness of the solutions, is as follows:

Let $f_1, \cdots, f_n$ be real-valued, continuous functions on the $(n+1)$-dimensional region

$$\text{(i)} \qquad 0 \leqq t \leqq a, \qquad -\infty < x_i < \infty \qquad (i = 1, \cdots, n).$$

Suppose that there exists a continuous function $L(r)$, $0 \leqq r < \infty$, satisfying on the region (i) *the n inequalities*

(ii). $$| f_i(t; x_1, \cdots, x_n) | < L(r),$$

where

(iii) $$r^2 = x_1^2 + \cdots + x_n^2,$$

and having the property that

(iv) $$\int^{\infty} dr/L(r) = \infty.$$

Then, if $c_1, \cdots, c_n$ is any set of integration constants, the differential equations

(v) $$x_i' = f_i(t; x_1, \cdots, x_n)$$

and the initial conditions

(vi) $$x_i(0) = c_i$$

have a solution $x_i = x_i(t)$ on the whole t-range, $0 \leqq t \leqq a$, admitted in (i).

Suppose that this assertion is false. Then there exists a positive number t_0 not exceeding the value of a and having the property that (v) and (vi) have on the interval $0 \leqq t < t_0$ a solution $x_i = x_i(t)$ which cannot be extended to the closed interval $0 \leqq t \leqq t_0$. Hence, if the local existence theorem of (v) and (vi) is applied in exactly the same way as above (where $n = 1$), it follows that there cannot exist an increasing sequence $t_1, t_2, \cdots$ tending to t_0 and having the property that each of the n sequences $x_i(t_1), x_i(t_2), \cdots$ is bounded. But the non-existence of such a sequence $t_1, t_2, \cdots$ means, by (iii), that

(vii) $$r = r(t) \to \infty \quad \text{as} \quad t \to t_0 - 0.$$

And it is easy to see that (vii) contradicts (ii) and (iv).

In fact, differentiation of (iii), where $r = r(t)$, gives

$$(rr')^2 \leqq (x_1^2 + \cdots + x_n^2)(x_1'^2 + \cdots + x_n'^2) = r^2(f_1^2 + \cdots + f_n^2),$$

by (v). Hence, from (ii),

$$dt \geqq | dr |/L(r),$$

if the constant factor $n^{1/2}$ is thought of as absorbed by the function sign L. But the last formula line implies that the inequality

$$\lim_{t \to t_0 - 0} t - c \geqq \lim_{R \to \infty} \int_C^R dr/L(r)$$

holds for a certain pair of integration constants c, C. Since the limit on the left has a finite value, this inequality contradicts (iv).

It is clear from the proof that the continuity of the positive function $L(r)$ can be replaced by more general local assumptions.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] E. Hilb, "Nichtlineare Differentialgleichungen," *Encyklopädie der mathematischen Wissenschaften*, vol. II B, art. 6 (1916), pp. 563-603.

[2] P. Painlevé, *Leçons sur la théorie analytique des équations différentielles professées à Stockholm* (1895), Paris, 1897.

[3] A. Wintner, "The non-local existence problem of ordinary differential equations," *American Journal of Mathematics*, vol. 67 (1945), pp. 277-284.

REDUCTION OF AN n-TH ORDER LINEAR DIFFERENTIAL EQUATION AND m-POINT BOUNDARY CONDITIONS TO AN EQUIVALENT MATRIX SYSTEM.*

By Randal H. Cole.

1. Introduction. The linear differential equation of order n,

$$u^{(n)} + P_1 u^{(n-1)} + \cdots + P_n u = 0, \tag{1.1}$$

with boundary conditions of various types has been extensively discussed by Birkhoff [1],[1] Wilder [2], Tamarkin [3], and others. Much attention has also been devoted to systems of first order equations,[2] and in particular to systems which may be written in the matrix [3] form,

$$\mathfrak{Y}' = \{\lambda \mathfrak{R} + \mathfrak{Q}\}\mathfrak{Y}. \tag{1.2}$$

Here, as before, the problems are largely characterized by the type of the adjoined boundary relations. The results obtained by Birkhoff and Langer [5] in 1923 and by Langer [6] in 1939 may be cited as examples of significant developments in this field.

Although (1.1) is readily reducible to a system of n first order linear equations, obvious reductions do not yield systems which are linear in the parameter. A reduction of (1.1) to (1.2) has been obtained by Wilder [7], but it is not carried to the point where the matrices $\mathfrak{R}$ and $\mathfrak{Q}$ are in a form which permits the immediate application of known results for expansion problems. Further, the reducibility of boundary conditions associated with (1.1) to an equivalent set associated with (1.2) has not been considered.

The present paper shows that (1.1) is reducible to a familiar and convenient form of system (1.2), and that a set of m-point boundary conditions applying to (1.1) admits of a corresponding reduction to a set of matrix boundary conditions applying to (1.2). The theory of differential matrix

* Received May 1, 1945.

[1] Numbers in brackets refer to the bibliography.

[2] A review of the literature associated with such systems has been given by W. M. Whyburn [4].

[3] Square matrices of order n will be designated by German capital letters and their components will be represented by the corresponding lower case italic letters, that is, $\mathfrak{B}(x) \equiv (b_{i,j}(x))$. Similarly the lower case German letters will be reserved exclusively for the representation of vectors.

equations is, therefore, made available for the extension and generalization of the classical boundary and expansion problems associated with the n-th order linear differential equation.

2. Reduction of the n-th order equation to matrix form. The equation to be considered is (1.1) for which it is assumed that

(a) *each coefficient P_k is a polynomial in λ of the form*

$$P_k = \sum_{l=0}^{k} P_{k,l}(x)\lambda^l,$$

with $P_{k,l}(x)$ free from λ and indefinitely differentiable,

(b) *the algebraic equation,*

$$r^n + P_{1,1}r^{n-1} + \cdots + P_{n,n} = 0,$$

has n distinct roots, $r_1(x), r_2(x), \cdots, r_n(x)$, for all values of x on a fundamental interval (a, b).

Under these assumptions there are known to exist [7] two indefinitely differentiable matrices, $\mathfrak{B}(x)$ and $\mathfrak{C}(x)$, free from λ and such that any matrix solution of

$$\mathfrak{Z}' = \{\lambda\mathfrak{B}(x) + \mathfrak{C}(x)\}\mathfrak{Z} \tag{2.1}$$

has solutions of (1.1) as the components of its first row. Further, $\mathfrak{B}(x)$ and $\mathfrak{C}(x)$ are such that

$$b_{i,j}(x) = 0, \qquad j > i,$$
$$b_{i,i}(x) = r_i(x),$$

and

$$c_{i,j}(x) = 0, \qquad j > i+1.$$
$$c_{i,i+1}(x) = 1.$$

A matrix of the form of $\mathfrak{B}(x)$ will henceforth be referred to as a lower triangular matrix and one of the form of $\mathfrak{C}(x)$ as a bordered lower triangular matrix.

In view of the form of $\mathfrak{B}(x)$ there exists a non-singular matrix $\mathfrak{T}(x)$ such that

$$\mathfrak{T}^{-1}(x)\mathfrak{B}(x)\mathfrak{T}(x) = \mathfrak{R}(x),$$

where [4]

$$\mathfrak{R}(x) = (\delta_{i,j} r_j(x)).$$

Since $\mathfrak{T}$ satisfies the relation $\mathfrak{B}\mathfrak{T} = \mathfrak{T}\mathfrak{R}$, it may be chosen of lower triangular form with 1's on the main diagonal. The matrix $\mathfrak{T}^{-1}$ will then be similarly constituted. With this choice of $\mathfrak{T}$ it may be verified that $\mathfrak{Q}(x)$, defined by

$$\mathfrak{Q}(x) = \mathfrak{T}^{-1}(x)\mathfrak{C}(x)\mathfrak{T}(x) - \mathfrak{T}^{-1}(x)\mathfrak{T}'(x),$$

is a bordered lower triangular matrix. For $\mathfrak{R}$ and $\mathfrak{Q}$ so defined, a solution of (2.1) yields, through the relation $\mathfrak{Z}(x, \lambda) = \mathfrak{T}(x)\mathfrak{Y}(x, \lambda)$, a solution of

$$\mathfrak{Y}' = \{\lambda\mathfrak{R}(x) + \mathfrak{Q}(x)\}\mathfrak{Y}. \tag{2.2}$$

In view of the form of $\mathfrak{T}$, it follows that any matrix solution of (2.2) will have solutions of (1.1) as components of its first row. Further, if any solution of (2.2) is non-singular, the components of its first row constitute a fundamental set of solutions of (1.1). This fact may be established by considering the relations, defined by (2.2), which the components of $\mathfrak{Y}$ must satisfy. The form of $\mathfrak{R}$ and $\mathfrak{Q}$ is such that these relations reduce to

$$y'_{i,j} = \lambda r_i y_{i,j} + \sum_{\nu=1}^{i} q_{i,\nu} y_{\nu,j} + y_{i+1,j}.$$

It is, therefore, clear that if the first i components of the j-th column of $\mathfrak{Y}$ are identically zero, the $(i+1)$-th component is also identically zero. Thus, if the first component of the j-th column is zero, all the components of that column are zero and $\mathfrak{Y}$ is singular. This implies that if $\mathfrak{Y}$ is non-singular, no component of its first row is zero. If $\mathfrak{C}$ is any non-singular matrix free from x, $\mathfrak{Y}\mathfrak{C}$ is also a non-singular solution and, by the same argument has no zero components in its first row. Hence, if the components of the first row of $\mathfrak{Y}$ were not linearly independent, an appropriate choice of $\mathfrak{C}$ would furnish a contradiction.

We have shown that the first row of $\mathfrak{Y}$ is made up of u's, but there is still the question of the relationship between the derivatives of these u's and the components of $\mathfrak{Y}$. With this question in view, let the square matrix $\mathfrak{K}(u_1, u_2, \cdots, u_n)$, associated with any n functions $u_1(x), u_2(x), \cdots, u_n(x)$, each possessing $(n-1)$ derivatives, be defined by

$$\mathfrak{K}(u_1, u_2, \cdots, u_n) = (u_j^{(i-1)}(x)).$$

[4] The symbol $\delta_{i,j}$ is used in the sense, $\delta_{i,j} = 0$, if $i \neq j$; $\delta_{i,j} = 1$, if $i = j$.

We shall show that there exists a non-singular matrix $\mathfrak{S}(x, \lambda)$, such that, if $\mathfrak{Y}(x, \lambda)$ is any matrix solution of (2.2),

$$\mathfrak{S}(x, \lambda)\mathfrak{Y}(x, \lambda) = \mathfrak{R}(y_{1,1}, \cdots, y_{1,n}), \tag{2.3}$$

where $y_{1,1}, \cdots, y_{1,n}$, being components of the first row of $\mathfrak{Y}(x, \lambda)$, are solutions of (1.1).

To derive $\mathfrak{S}$, let the matrix $\mathfrak{S}^{(\eta)}$ be defined by the recurrence relation

$$\mathfrak{S}^{(\eta)} = d\mathfrak{S}^{(\eta-1)}/dx + \mathfrak{S}^{(\eta-1)}\mathfrak{S}^{(1)},$$

with $\mathfrak{S}^{(0)} = (\delta_{i,j})$ and $\mathfrak{S}^{(1)} = \{\lambda\mathfrak{R}(x) + \mathfrak{Q}(x)\}$. The η-th derivative of any matrix solution $\mathfrak{Y}$ is, therefore, given by

$$\mathfrak{Y}^{(\eta)} = \mathfrak{S}^{(\eta)}\mathfrak{Y}, \qquad (\eta = 0, 1, 2, \cdots).$$

The first row of $\mathfrak{Y}^{(\eta-1)}$ is clearly the η-th row of $\mathfrak{R}(y_{1,1}, \cdots, y_{1,n})$ so that

$$\mathfrak{R}(y_{1,1}, \cdots, y_{1,n}) = \sum_{\eta=1}^{n} (\delta_{i,\eta}\delta_{1,j})\mathfrak{S}^{(\eta-1)}\mathfrak{Y}.$$

The relation (2.3) follows immediately with $\mathfrak{S}(x, \lambda)$ defined by

$$\mathfrak{S}(x, \lambda) = \sum_{\eta=1}^{n} (\delta_{i,\eta}\delta_{1,j})\mathfrak{S}^{(\eta-1)}. \tag{2.4}$$

When so defined, $\mathfrak{S}(x, \lambda)$ is readily seen to have components which are polynomials in λ. We may show that it is non-singular by assuming that the components of $\mathfrak{S}^{(\eta)}$ are such that

$$s_{i,j}{}^{(\eta)} \equiv 0 \text{ if } j > i + \eta; \qquad s^{(\eta)}_{i,i+\eta} \equiv 1.$$

The component in the i-th row and j-th column of $\mathfrak{S}^{(\eta+1)}$ is therefore given by

$$s_{i,j}{}^{(\eta+1)} = ds_{i,j}{}^{(\eta)}/dx + \sum_{\nu=j-1}^{i+\eta} s_{i,\nu}{}^{(\eta)}s_{\nu,j}{}^{(1)},$$

which implies that

$$s_{i,j}{}^{(\eta+1)} \equiv 0 \text{ if } j > i + \eta + 1; \qquad s^{(\eta+1)}_{i,i+\eta+1} \equiv 1.$$

That is, the matrix $\mathfrak{S}^{(\eta+1)}$ is also of the assumed form and, since $\mathfrak{S}^{(0)}$ and $\mathfrak{S}^{(1)}$ are known to be of this form, the validity of the assumption is established by induction. In view of (2.4), therefore, $\mathfrak{S}(x, \lambda)$ is a lower triangular matrix with unit components on the main diagonal and is non-singular.

3. Boundary conditions. The boundary conditions to be adjoined to (1.1) apply at the end points of the fundamental interval and also at an

arbitrary set of intermediate points, $a_2, a_3, \cdots, a_{m-1}$. If the end points a and b, are denoted by a_1 and a_m, respectively, and if the designation is such that $a_\mu < a_{\mu+1}$, $(\mu = 1, 2, \cdots, m-1)$, these conditions are given by

$$(3.1) \quad V_i^{(1)}(u) + V_i^{(2)}(u) + \cdots + V_i^{(m)}(u) = 0, \quad (i = 1, 2, \cdots, n),$$

where

$$V_i^{(\mu)}(u) = v_{i,1}^{(\mu)} u(a_\mu, \lambda) + v_{i,2}^{(\mu)} u'(a_\mu, \lambda) + \cdots + v_{i,n}^{(\mu)} u^{(n-1)}(a_\mu, \lambda)$$

with coefficients, $v_{i,j}^{(\mu)}$, analytic functions of λ. If $\mathfrak{k}(x, \lambda)$ is defined to be the vector whose components are $u(x, \lambda)$ and its first $(n-1)$ derivatives, these relations may be written in the matrix form,

$$\sum_{\mu=1}^{m} \mathfrak{V}^{(\mu)}(\lambda) \mathfrak{k}(a_\mu, \lambda) = \mathfrak{o},$$

with

$$\mathfrak{V}^{(\mu)}(\lambda) = (v_{i,j}^{(\mu)}(\lambda)).$$

Consider the system

$$(3.2\text{ a}) \qquad \mathfrak{y}' = \{\lambda \mathfrak{R}(x) + \mathfrak{Q}(x)\} \mathfrak{y}$$

$$(3.2\text{ b}) \qquad \sum_{\mu=1}^{m} \mathfrak{W}^{(\mu)}(\lambda) \mathfrak{y}(a_\mu, \lambda) = \mathfrak{o},$$

where $\mathfrak{W}^{(\mu)}(\lambda)$ is defined as a matrix of analytic functions of λ by

$$\mathfrak{W}^{(\mu)}(\lambda) = \mathfrak{V}^{(\mu)}(\lambda) \mathfrak{S}(a_\mu, \lambda).$$

The equation (3.2 a) is the vector equation corresponding to (2.2) and has the general solution $\mathfrak{Y}(x, \lambda) \mathfrak{c}(\lambda)$, where $\mathfrak{Y}(x, \lambda)$ is any non-singular solution of (2.2) and $\mathfrak{c}(\lambda)$ is an arbitrary vector free from x. The first component of any solution of (3.2) is clearly a solution of (1.1) and (3.1), and conversely, any solution of the latter system will yield, under appropriate transformations, a solution of (3.2).

The results obtained by Langer [6] for a system in the complex domain represent significant developments when specialized to the present problem. It is, therefore, worthy of note that the substitution (cf. [6, p. 154])

$$\mathfrak{y}(x, \lambda) = (\delta_{i,j} e^{\phi_j(x)}) \mathfrak{u}(x, \lambda),$$

where $\phi'_j(x) = q_{j,j}(x)$, reduces (3.2) to Langer's system. The substitution has the effect of replacing (3.2) by an equivalent and similar system in which

$\mathfrak{R}(x)$ is unchanged and $\mathfrak{Q}(x)$ is replaced by a matrix whose main diagonal components are identically zero.

UNIVERSITY OF WESTERN ONTARIO.

BIBLIOGRAPHY.

1. G. D. Birkhoff, "Boundary value and expansion problems of ordinary linear differential equations," *Transactions of the American Mathematical Society*, vol. 9 (1908), pp. 373-395.

2. C. E. Wilder, "Expansion problems of ordinary linear differential equations with auxiliary conditions at more than two points," *Transactions of the American Mathematical Society*, vol. 18 (1917), pp. 415-442.

3. J. D. Tamarkin, *On Certain General Problems of the Theory of Ordinary Linear Differential Equations and the Expansion of an Arbitrary Function in Series*, Petrograd, 1917.

4. W. M. Whyburn, "Differential equations with general boundary conditions," *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 692-704.

5. G. D. Birkhoff and R. E. Langer, "The boundary problems and developments associated with a system of ordinary linear differential equations of the first order," *Proceedings of the American Academy of Arts and Sciences*, vol. 58 (1923), pp. 51-128.

6. R. E. Langer, "The boundary problem of an ordinary linear differential system in the complex domain," *Transactions of the American Mathematical Society*, vol. 46 (1939), pp. 151-190.

7. C. E. Wilder, "Reduction of the ordinary linear differential equation of the n-th order whose coefficients are certain polynomials in a parameter to a system of n first order equations which are linear in the parameter," *Transactions of the American Mathematical Society*, vol. 29 (1927), pp. 497-506.

LINEAR VARIATIONS OF CONSTANTS.*

By Aurel Wintner.

1. Let $D = D_z$ be an homogeneous, linear differential operator of order n, with coefficients which are single-valued and regular near a point, say the origin, of the z-plane. Then every solution $w = w(z)$ of $Dw = 0$ is a linear combination of a finite number of functions each of which is of the form $z^\lambda (\log z)^l L(z)$, where λ is one of the "characteristic exponents," l a non-negative integer not exceeding $n - 1$, and $L(z)$ a Laurent series convergent near $z = 0$. This is a classical result of Fuchs (1866), who also determined the conditions under which no $L(z)$ has an essential singularity at $z = 0$, that is, under which the expansion (near $z = 0$) of every solution $w(z)$ can be calculated recursively (rather than only by using infinite determinants or equivalent processes). These conditions, subsequently found in Riemann's posthumous notes also, prove to be fundamental, since they do not involve the knowledge of any solution: Necessary and sufficient is that, when the coefficient of the n-th derivative of w in $Dz = 0$ is 1, the singularity of the coefficient of the k-th derivative be a pole of order $n - k$ (at most), where $k = 0, 1, \cdots$.

The necessity of this criterion, though essential in the hypergeometric theory and its generalizations (Riemann; Pochhammer), is quite on the surface. In addition, it is of an accidental nature, in the sense that it ceases to hold when $Dw = 0$ is generalized to a system of n linear differential equations of the first order, $w' = F(z)w$, where $w = w(z)$ now is a vector with n components and $F(z)$ a matrix of n times n functions.

On the other hand, the sufficiency of the criterion can be transferred to this, more general and symmetric, case without any additional trouble. In fact, the sufficiency of the criterion then states simply that all Laurent series occurring in the general solution of $w' = F(z)w$ are free of essential singularities whenever the matrix $F(z)$ has at $z = 0$ a simple pole (at most).

All the known proofs of this criterion (as well as of its particular case belonging to $Dw = 0$) are arduous. They are more or less straightforward variants of two main types. The *first* type of proof is substantially that of Fuchs. This proof considers, first under the mere assumption that $F(z)$ is single-valued and regular near $z = 0$, the local monodromy group, consisting of the substitutions to which the solution vectors $w = w(z)$ are subjected by

* Received October 23, 1945.

a circuit about $z=0$. It then puts the monodromy matrices into their (common) Jordan normal form. From this, it is possible to conclude that all solutions are linear combinations of a finite number of functions each of which is composed of three factors z^λ, $(\log z)^l$, $L(z)$. Finally, it is shown (and this can be effected in various ways) that no $L(z)$ can have an essential singularity if the singularity of the coefficient matrix is a simple pole.

In contrast, the *second* type of proof (which, in the particular case of $Dw=0$, goes back to Frobenius) confines itself to the case of a simple pole from the beginning. It consists in proving, by the method of undetermined coefficients, that the expansions of the solutions to be supplied by the answer all exist and form n, and not less than n, linearly independent solution vectors. And this requires, besides a convergence proof, a counting of constants, which becomes elaborate not only when one of the above integers l becomes distinct from 0 (that is, when the matrices of the monodromy group have a multiple invariant factor) but also when the difference of two of the (normalized) characteristic exponents λ happens to be a real integer. What makes matters worse in this formal theory is that, just as in the theory of the hypergeometric equation, the case of (ostensibly) exceptional λ-sets has no function-theoretical significance, since the monodromy group determines just the residue class (mod 1) of an exponent λ, rather than λ itself.

Both of these proofs (the "Riemannian" and the "Weierstrassian") apply the full force of the theory of the elementary divisors; the first, via the normal form of the monodromy group, the second, as a counting machine. But a glance at the explicit form of the final result shows that what is actually proved can finally be worded so as to involve neither the characteristic numbers ($=e^{\lambda/2\pi i}$) nor the elementary divisors of the monodromy group; namely, as follows:

(i) *If all n^2 elements of the coefficient matrix, $F=F(z)$, of $w'=F(z)w$ are regular in a circle $|z|<a$ except for simple poles at $z=0$, and if $R=R_F$ denotes the residue $(zF(z))_{z=0}$, then one solution matrix is a matrix product of the form $P(z)z^R$, where $P(z)$ is a matrix regular at $z=0$ (that is, $P(z)=P(0)+P'(0)z+\frac{1}{2}P''(0)z^2+\cdots$, if $|z|<a$) and* $\det P(z)\neq 0$, *if $0<|z|<a$.*

It is understood that by a "solution matrix" is meant any matrix the columns of which are solution vectors (cf. (12) below), and that $\det P(z)\neq 0$ makes the solution matrix in question a "fundamental matrix."

The matrix multiplying $P(z)$ is defined, of course, by $z^R=e^{R\log z}$, where e^A denotes the matrix

$$e^A=\sum_{m=0}^{\infty}A^m/m!$$

and $\log z$ is thought of on its Riemann surface. Since e^A has the period $2\pi iE$, this agrees with the fact that only $e^{\lambda/2\pi i}$, that is, only the residue class (mod 1) of a characteristic exponent λ, is defined by the monodromy group.

If Jordan's normal form, say J, of A is known, it can be used in order to "sum" the infinite series of e^A. In fact, if $A = TJT^{-1}$, then $e^A = Te^JT^{-1}$, and e^J can be "summed" by using, for the powers of J, the recursion formula supplied by Cayley's theorem, $f(B) = 0$, where $f(s) = \det(sE - B)$. But the significance of such "summations" is secondary indeed, and so the truly function-theoretical wording of the classical result, a wording based on the exponential function z^R rather than on the counting machine of the elementary divisors of R, is just (i).

An *experimentum crucis* is supplied by infinite matrices R which are bounded in Hilbert's sense. Then z^R is defined as before and is a non-singular, bounded matrix (at every $z \neq 0$). But now there is no Cayley theorem available for "summation" purposes and, what is much worse, no analogue of the Jordan-Weierstrass theory can exist (Toeplitz), not even in the apparently harmless case of a completely continuous R.

2. Let the dotted circle, $0 < |z| < a$, be replaced by an interval, $0 < t < a$, and let the pole of the coefficient matrix be split off by placing $F(t) = t^{-1}T + G(t)$ in (i). Then the "exponential" wording of the classical theorem suggests the possibility of an extension to the case in which the deviation, $G(t)$, of $F(t)$ from the principal part, $t^{-1}R$, is not a regular power series but a function restricted by "real" smoothness assumptions only. Neither of the classical types of proof, sketched above, is then available (the second not, because there are no coefficients to be compared, and the first not, because there are no "circuits," hence no monodromy matrices, in the one-dimensional case). Nevertheless, the extensions in question prove to exist, at least under the assumption that the constant matrix R, representing the formal residue of $F(t)$, is "small" enough; the limitation of its "size" depending only on the dimension number, n (rather than on the choice of $G(t) = t^{-1}T - F(t)$ also).

In order to define this notion, let $|A|$ denote the greatest among the n^2 absolute values $|a_{ik}|$, if (a_{ik}) is the matrix A. Then it is clear that

$$(1) \qquad |AB| \leqq n|A|\,|B| \quad \text{and} \quad |A + B| \leqq |A| + |B|.$$

Hence, if $[X] = [X]_R$ denotes the matrix

$$(2) \qquad [X] = RX - XR,$$

there exists a number r satisfying the inequality

$$| [X] | \leqq r | X |, \tag{3}$$

where the matrix X is arbitrary, the matrix R is fixed, and the factor $r = r_R$ is independent of X. Let r be the least factor satisfying (3) for every X. This unique $r = r_R$ will be called the *cross-modulus of the matrix* R. According to (1) and (2), the cross-modulus is subject to the inequality

$$r \leqq 2n | R |, \tag{4}$$

no matter what R may be.

By using the appropriate metric, it is possible to extend the definition of $| A |$, and therefore that of the cross-modulus, to the case of Hilbert's space (and other linear spaces). The following proofs then require nothing but the customary transcription.

A "size" of the formal residue being defined by its cross-modulus, the theorem announced above can now be formulated as follows:

(i_0) *Let R be a constant matrix the cross-modulus of which is less than 1, and let $G(t)$ be a matrix function which is continuous on an half-open interval, $0 < t \leqq t_0$, and remains bounded,*

$$G(t) = O(1), \quad \text{as} \quad t \to +0. \tag{5}$$

Then one solution matrix of

$$x' = F(t)x, \quad \text{where} \quad F(t) = R/t + G(t), \tag{6}$$

is a matrix product of the form

$$P(t)t^R, \qquad (t^R = e^{R \log t}, \quad \log t \to -\infty), \tag{7}$$

where

$$P(+0) \quad \text{exists} \tag{8}$$

(as a finite limit) and

$$\det P(+0) \neq 0. \tag{9}$$

[*Actually,* (8) *can be improved to*

$$P(t) = P(+0) + O(t), \tag{10}$$

as $t \to +0$.]

If (i_0) did not restrict the residue of the principal part of F, it would represent a complete dual of (i) for the non-analytic case. What concerns the secondary part of F, the assumption (5) is sure to be satisfied if $G(t)$ is uniformly continuous. But (5) does not assume this, that is, $G(+0)$ need not exist.

The proof of (i_0) will depend on a simple "Abelian" lemma (in contrast to some "Tauberian" consideration).

3. Let $X, U, A, \cdots$ denote matrices (with n rows and n columns), and let a' denote da/dt, whether $a = X$ or $a = x$, where t is a real variable.

If $A(t)$ is a continuous function on an interval, then

$$x' = A(t)x \tag{11}$$

has on the whole interval a unique solution $x = x(t)$ which, at a point t_0 of the interval, becomes an arbitrary initial vector, $x(t_0)$. If $X = X(t)$ denotes the matrix the columns of which are n solutions, say $x = x^1, \cdots, x = x^n$, then, since

$$X = (x^1, \cdots, x^n), \tag{12}$$

(11) is equivalent to

$$X' = A(t)X, \tag{13}$$

and the determination of the general solution of (11) is equivalent to the determination of a solution of (13) in which the columns are linearly independent, that is,

$$\det X(t) \neq 0. \tag{14}$$

Since, if $\operatorname{tr} A$ denotes the sum of the diagonal elements of A,

$$\det X(t) = \det X(t_0) \cdot \exp \int_{t_0}^{t} \operatorname{tr} A(s)\, ds \tag{15}$$

is an identity in t by virtue of (13) alone (Jacobi), it is needless to specify whether the linear independence of the columns of (12) be meant for some t or for every t. Clearly, (13) remains true if X is replaced by XC (but not by CX), where C is any constant matrix. Hence, if a solution of (13) satisfying (14) is called a fundamental matrix of (11), a matrix is a fundamental matrix of (11) if and only if it is a product $X(t)C$, where $X(t)$ is some fundamental matrix and C denotes a constant matrix of non-vanishing determinant.

Instead of considering, as before, just one system of homogeneous linear differential equations, consider two of them, (11) and

(16) $$y' = B(t)y$$

or, equivalently, (13) and

(17) $$Y' = B(t)Y,$$

where A and B are arbitrary continuous functions on a (common) t-interval. If (11) is to be transformed into (16) by the procedure of the variation of constants (Lagrange), then, since $X(t)$ had to be placed in front of the constant matrix C, the "varied" form, say $U = U(t)$, of $C =$ const. must be set up as follows:

(18) $$Y = XU.$$

And this leads to the

Rule for the Variation of Constants from the Right. *If $X(t)$ is a fixed fundamental matrix of* (11), *all the fundamental matrices $Y(t)$ of* (16), *and only these matrices, are represented by the product* (18) *in which $U(t)$ denotes an arbitrary fundamental matrix of $u' = K(t)u$, that is, any solution $U = U(t)$ of*

(19) $$U' = K(t)U$$

satisfying

(19 bis) $$\det U(t) \neq 0,$$

where the coefficient matrix, $K = K(t)$, is the ordinary transform of the perturbation $B(t) - A(t)$, that is,

(20) $$K = X^{-1}(B - A)X.$$

This follows by direct substitutions. In fact, if U is thought of as defined by (18) (which is possible, since $\det X \neq 0$), differentiation of (18) gives

$$Y' = X'U + XU'.$$

Hence, if Y', Y, X' are substituted from (17), (18), (13) respectively,

$$BXU = AXU + XU'.$$

But this can be written in the form (19), if $K = K(t)$ is defined by (20).

If (18) is replaced by

$$Y = VX, \tag{21}$$

what results is the

Rule for the Variation of Constants from the Left. *If $X(t)$ is a fundamental matrix of* (11), *all fundamental matrices $Y(t)$ of* (16), *and only these matrices, are represented by the product* (21) *in which $V = V(t)$ denotes any solution of*

$$V' = B(t)V - VA(t) \tag{22}$$

satisfying

$$\det V(t) \neq 0. \tag{22 bis}$$

The verification proceeds in the same way as before. But the two rules are quite different in structure. First, (19) is equivalent to $u' = K(t)u$, a homogeneous, linear system of order n, whereas the "unsymmetric bracket" condition for the elements of the matrix V represents a homogeneous, linear system of order n^2. Next, the initial fundamental matrix $X(t)$ occurs, via (20), in (19) but does not enter into (22). Finally, the second rule is not a true process of variation of constants, since it is based on (21), where, if V is replaced by a constant matrix, C, the product CX does *not* become (in general) a solution of (13).

To what the second rule actually corresponds is a local formulation of the Riemann-Fuchs equivalence problem of species ($=$ $\mathfrak{A}$rt $=$ Poincaré's *espèce*). If this is compared with Schlesinger's theory of simply canonical systems,* then, since the prototype of the problem of (i) is Cauchy's elementary system $zw' = Rw$ (a system possessing the solution matrix z^R), it is clear that anything pertaining to (i) should be based on (21), (22), rather than on (18), (19), (20), the method of the variation of constants in its strict sense.

4. Since the substitution

$$t \ (\rightarrow)\ e^{-t}, \qquad (\lim_{t\rightarrow +0} \rightarrow \lim_{t\rightarrow\infty}), \tag{23}$$

transforms the prototype, $tx' = Rx$, of (6) into $x' = Rx$, where $R = \text{Const.}$, it is convenient to transform (6) itself by (23). Then (6) becomes

$$-e^t x' = F(e^{-t})x, \quad \text{where} \quad F(e^{-t}) = e^t F + G(e^{-t}),$$

---

* L. Schlesinger, "Ueber eine Klasse von Differentialsystemen beliebiger Ordnung mit festen kritischen Punkten," *Crelle's Journal für die reine und angewandte Mathematik*, vol. 141 (1912), pp. 96-145, where further references will be found.

that is,

$$x' = A(t)x, \quad \text{where} \quad A(t) = -R - e^{-t}G(e^{-t}).$$

Since (11) is equivalent to (13) and (12), this can be written in the form

$$X' = -RX + H(t)X, \tag{24}$$

where

$$H(t) = -e^{-t}G(e^{-t}), \qquad (t \to \infty). \tag{25}$$

According to (25), the assumption (5) is equivalent to

$$H(t) = O(e^{-t}) \quad \text{as} \quad t \to \infty. \tag{26}$$

In particular, the prototype of (25) is $X' = -RX$, where $-R = \text{Const.}$ Since $X = e^{-tR}$ is a solution of this prototype, and since the *second* of the rules of the variation of constants should be applied, what suggests itself for (24) is the substitution

$$X = Ze^{-tR} \tag{27}$$

(and *not* $X = e^{-tR}Z$, which would correspond to the first rule). But, since

$$(Ze^{-tR})' = (Z' - ZR)e^{-tR},$$

the substitution (27) transforms (24) into

$$Z' - ZR = -RZ + H(t)Z;$$

and here the principal term is the Poissonian bracket, (2), of Z:

$$Z' = -[Z] + H(t)Z, \quad \text{where} \quad [Z] = RZ - ZR. \tag{28}$$

The formal properties of the bracket will now make possible a very easy proof for a general "Abelian" limit theorem.

5. The lemma in question is as follows:

(ii_0) *If the cross-modulus of the constant matrix R is less than* 1, *and if the matrix function $H(t)$ is continuous on an half-line*, $t_0 \leqq t < \infty$, *and satisfies* (26), *then* (28) *has a (unique) solution $Z = Z(t)$ satisfying*

$$Z(t) \to E \quad \text{as} \quad t \to \infty. \tag{29}$$

[*Actually*, (29) *can be improved to*

(29 bis) $$Z(t) = E + O(e^{-t}),$$

which, in view of (28) *and* (26), *implies that* $Z'(t) = O(e^{-t})$.]

Here E denotes the unit matrix. In reality, what concerns the rôle of (ii$_0$), an arbitrary $C = \text{Const.}$ of non-vanishing determinant, rather than just $C = E$, is needed in the boundary condition (29). But the normalization $C = E$ is essential in the following proof of (ii$_0$) itself, since the non-commutative nature of the operations connected with (2) prevents the use of an arbitrary C.

The standard form of the successive approximations to (28) and (29) is

(30) $$Z_{m+1}(t) = E + \int_t^\infty [Z_m(s)]ds - \int_t^\infty H(s)Z_m(s)ds,$$

where

(31) $$Z_0(t) = E.$$

The convergence of the integrals (30) will have to be ascertained, of course.

The point in choosing precisely E in (29) is that $X = E$ is the only matrix ($\neq 0$) for which (2) vanishes when R is unspecified. Thus, from (31) and (30),

$$Z_1(t) = E - \int_t^\infty H(s)ds.$$

This means that

(32) $$Y_1(t) = - \int_t^\infty H(s)ds,$$

if $Y_{m+1}(t)$ denotes the difference

(33) $$Y_{m+1}(t) = Z_m(t) - Z_{m-1}(t).$$

But it is clear from (33) that (30) is equivalent to

(34) $$Y_{m+1}(t) = \int_t^\infty [Y_m(s)]ds - \int_t^\infty H(s)Y_m(s)ds,$$

since $[A - B] = [A] - [B]$, by (2). Furthermore, (32) and (26) assure the existence of $Y_1(t)$. Consequently, the integrals (30) will be proved to be convergent if the same is proved of the integrals (34).

Let r denote the cross-modulus of R. Then, according to (3), (1) and (34),

$$| Y_{m+1}(t) | \leqq r \int_t^\infty | Y_m(s) | \, ds + 2n \int_t^\infty | H(s) | \; | Y_m(s) | \, ds.$$

Hence,

$$\beta_{m+1}(t) \leqq r \int_t^\infty \beta_m(s) \, ds + 2n\beta_m(t) \int_t^\infty | H(s) | \, ds,$$

where

(35) $$\beta_m(t) = \underset{t \leqq s < \infty}{\text{l. u. b.}} | Y_m(s) |.$$

For the present, $\beta_m(t) = \infty$ is not precluded. But, by the assumption (26),

(36) $$| H(t) | < pe^{-t},$$

where p is a positive constant. Hence, by the inequality preceding (35),

(37_m) $$\beta_{m+1}(t) < r \int_t^\infty \beta_m(s) \, ds + 2np\beta_m(t) e^{-t},$$

where, according to (32) and (35),

(37_0) $$\beta_1(t) \leqq \int_t^\infty | H(s) | \, ds < pe^{-t}.$$

Thus, from (37_1),

$$\beta_2(t) < rpe^{-t} + 2np^2 e^{-2t}.$$

Hence, if t_ϵ is so large that

(38) $$2npe^{-t} < \epsilon \quad \text{when} \quad t_\epsilon < t < \infty,$$

where $\epsilon > 0$ is arbitrary,

$$\beta_2(t) < p(r + \epsilon) e^{-t}.$$

Consequently, from (37_2),

$$\beta_3(t) < p(r + \epsilon)(re^{-t} + 2npe^{-2t})$$

and so, if t satisfies (38),

$$\beta_3(t) < p(r + \epsilon)^2 e^{-t}.$$

It is now clear that (37_m) and (38) lead to

$$\beta_{m+1}(t) < p(r + \epsilon)^m e^{-t},$$

where $m = 1, 2, \cdots$ and $t_\epsilon < t < \infty$.

Accordingly, from (35) and (33),

$$(39)\qquad |Z_m(t)-Z_{m-1}(t)|<p\theta^{m+1}e^{-t}\quad\text{if}\quad t^0<t<\infty,$$

if θ denotes the sum $r+\epsilon$ and t^0 is the t_ϵ belonging to a fixed $\theta=\theta_\epsilon$.

Let ϵ be chosen so as to make $\theta=r+\epsilon$ less than 1 (this is possible, since r, the cross-modulus of R, is supposed to be less than 1). Then (39) and (31) show that the limit process

$$Z_m(t)\to Z(t),\qquad\qquad m\to\infty$$

defines, by uniform convergence, a (continuous) function $Z(t)$ which satisfies the relation (29 bis) and, in view of (36) and (2), the limiting case of the recursion formula (30), that is, the identity

$$Z(t)=E+\int_t^\infty [Z(s)]ds-\int_t^\infty H(s)Z(s)ds.$$

Finally, since $Z(t)$ and $H(t)$ are continuous, this identity implies that $Z(t)$ has a (continuous) derivative and is a solution of (28).

This proves (ii_0), except for the parenthetical assertion claiming the uniqueness of $Z(t)$. But this follows by applying the successive approximations to the difference of two (allegedly distinct) solutions in the usual manner.

In order to deduce (i_0) from (ii_0), let X be defined by (12). Then the substitutions (23), (25), (27) transform (6) and (5) into (28) and (26) respectively. Hence, if the assumptions of (i_0) are satisfied, it follows from (ii_0) that (6) has a fundamental matrix, $X(t)$, which is of the form (7), where, in view of (23) and (29 bis),

$$P(t)=E+O(t)\quad\text{as}\quad t\to+0.$$

Accordingly, (10) and (9) are satisfied, the limit (8) being the unit matrix.

6. If R in (6), (7) is the zero matrix, then (8), (9) remain true if (5) is relaxed to

$$\int_{+0}^{\cdot}|G(t)|\,dt<\infty.$$

This known Abelian fact can be formulated as follows:

(iii_0) *If $A(t)$ is continuous on an half-open interval $0<t\leqq t_0$ and behaves, as $t\to+0$, so as to become absolutely integrable, and if $X(t)$ denotes a fundamental matrix of $x'=A(t)x$, then $X(+0)$ exists (as a finite limit).*

The proof is the same as, though of course simpler than, that of (ii_0). In fact, if the interval $0 < t \leqq t_0$ is short enough, then, by the assumptions of (iii_0),

$$n \int_{t}^{t_0} |A(t)|\,dt \leqq \tfrac{1}{2} \tag{40}$$

(n denotes the dimension number). In view of the principle of superposition, it is sufficient to prove (iii_0) for the fundamental matrix, $X(t)$, which is assigned by the initial condition $X(t_0) = E$. Then, according to the formulation (13) of (11), the successive approximations are defined by

$$X_{m+1}(t) = E - \int_{t}^{t_0} A(s) X_m(s)\,ds, \tag{41}$$

where $X_0(t) = E$; hence $X_1(t) = E - \int_{t}^{t_0} A(s)\,ds$. But, if

$$\alpha_m(t) = \underset{t \leqq s \leqq t_0}{\mathrm{l.\,u.\,b.}} \,|X_m(s) - X_{m-1}(s)|, \tag{42}$$

(so that, for the present, $\alpha_m(t) = \infty$ is allowed), it is clear from (41) and (1) that

$$\alpha_{m+1}(t) \leqq \int_{t}^{t_0} |A(s)|\, n\alpha_m(s)\,ds \leqq n\alpha_m(t) \int_{t}^{t_0} |A(s)|\,ds.$$

It follows, therefore, from (40) that $\alpha_{m+1}(t) \leqq \frac{1}{2}\alpha_m(t)$; hence $\alpha_{m+1}(t) \leqq (\frac{1}{2})^m \alpha_1(t)$. Finally, the difference of the matrices $X_0(t)$, $X_1(t)$, which were given after (41), is majorized by the integral (40), and so $\alpha_1(t) \leqq \frac{1}{2}$, by the case $m = 1$ of the definition (42). Consequently, $\alpha_m(t) \leqq (\frac{1}{2})^m$. In view of (42), this implies that the convergence of the successive approximations is uniform (on the half-open interval $0 < t \leqq t_0$). Since this, in turn, implies that the limit function is uniformly continuous (the functions (41) being continuous), the proof of (iii_0) is complete.

(iii_0 bis). *Under the assumptions of* (iii_0),

$$\det X(+0) \neq 0 \tag{43}$$

(*if $X(t)$ is a fundamental matrix*).

This is a corollary of (iii_0). In fact, if (43) were false, it would follow from (15) and (14), where $0 < t \leqq t_0$, that the integral on the right of (15)

is not $O(1)$ as $t \to +0$. But this is excluded by the (absolute) integrability of $\dot{A}(t)$, which is assumed in (iii_0).

Another corollary of (iii_0) is the following theorem:

(iii_m) *If $A(t)$ has on an half-line $m(\geqq 0)$ continuous derivatives satisfying*

$$(44) \qquad \int^{\infty} |A(t)|\, dt < \infty, \quad \int^{\infty} t\,|A'(t)|\, dt < \infty, \cdots,$$
$$\int^{\infty} t^m\, |A^{(m)}(t)|\, dt < \infty,$$

then every solution vector of $x = x(t)$ of $x' = A(t)x$ possesses an asymptotic representation of the form

$$(45) \qquad x(t) = \sum_{k=0}^{m} c_k t^{-k} + o(t^{-m}) \quad \text{as} \quad t \to \infty.$$

Moreover, if c_0 is any constant vector, there exists one and only one solution $x(t) = x(t; c_0)$ satisfying (45), where

$$c_1 = c_1(c_0),\ c_2 = c_2(c_0), \cdots,\ c_m = c_m(c_0).$$

First, the substitution

$$(46) \qquad t \to t^{-1}, \qquad (\lim_{t\to\infty} \to \lim_{t\to+0})$$

transforms $x' = A(t)$ into $x' = {}^*A(t)x$, where, if ${}^*A(t)$ is denoted just by $A(t)$,

$$A(t)\,dt \to -A(t)\,dt$$

holds in virtue of (46). Hence, in virtue of (46),

$$\Big(\int^{\infty} |A(t)|\, dt < \infty\Big) \to \Big(\int_{+0} |A(t)|\, dt < \infty\Big).$$

But the existence of $x(+0)$ means the existence of some c satisfying $x(t) = c + o(1)$ as $t \to +0$. Consequently, the case $m = 0$ of (iii_m) is equivalent to (iii_0) and (iii_0 bis) together.

Next, let $A(t)$ have a first derivative on an interval $0 < t \leqq t_0$. Then

$$x'' = A'(t)x + A(t)x', \quad \text{since} \quad x' = A(t)x.$$

But what is in this formula line can simply be written as $x' = A(t)x$, if

$$n, \qquad x, \qquad A(t)$$

are replaced by

$$2n, \qquad \begin{pmatrix} x \\ x' \end{pmatrix}, \qquad \begin{pmatrix} A(t) & 0 \\ A'(t) & A(t) \end{pmatrix}$$

respectively, where n is the dimension number and 0 denotes the n-rowed zero matrix. In particular, the new $A(t)$ satisfies the conditions of (iii_0) if and only if the old $A(t)$ has a continuous first derivative and

$$\int_{+0} |A(t)|\, dt < \infty, \qquad \int_{+0} |A'(t)|\, dt < \infty.$$

[Actually, the latter two conditions are tautological, since the convergence of the second integral implies that of the first and, as a matter of fact, the existence of a finite limit for $A(t)$ as $t \to +0$.]

Hence, if the conditions just mentioned are satisfied, it follows from (iii_0) that the old $x(t)$ and its derivative have finite limits, $x(+0)$ and $x'(+0)$. Consequently, if $x(t)$, where $0 < t \leqq t_0$, is defined to be $x(+0)$ at $t = 0$, then, by a standard theorem of differential calculus, $x(t)$ will have a continuous first derivative on the closed interval $0 \leqq t \leqq t_0$. Accordingly, an application of Taylor's rule gives $x(t) = x(0) + x'(0)t + o(t)$ as $t \to +0$.

Finally, it is clear from the remarks following (46) that, in virtue of the transformation (46), the case $m = 1$ of the $m - 1$ conditions (44) is equivalent to the pair of conditions represented by the last formula line. This proves (iii_m) for the case $m = 1$. And the case of an arbitrary m contains nothing new, since the above processes of differentiation can of course be repeated.

Remark. It is clear from this reduction of (iii_m) to (iii_0) that (ii_0) implies a theorem, say (ii_m), which corresponds to (iii_m) in the same way as (ii_0) to (iii_0). However, the condition $r < 1$, the first assumption of (ii_0), must then be replaced by $r < 1/(m+1)$. In fact, if $m = 1$, then what correspond to the dimension number n and to the matrices

$$R, \qquad Z, \qquad H,$$

of (ii_0) become, in (ii_1), the dimension number $2n$ and the matrices

$$\begin{pmatrix} R & 0 \\ 0 & R \end{pmatrix}, \qquad \begin{pmatrix} RZ - ZR & 0 \\ 0 & RZ - ZR \end{pmatrix}, \qquad \begin{pmatrix} H & 0 \\ H' & H \end{pmatrix}$$

respectively. Hence, if r denotes the cross-modulus of the old R, all that is clear is that the cross-molulus of the new R cannot exceed $2r$. Thus, the limi-

tation which, either just by the proof or by the true nature of (ii_m), is imposed on the cross-modulus of R becomes with increasing m so severe that no R distinct from the $R=0$ of (iii_∞) remains admitted in (ii_∞).

7. The proof of (i_0) depended on an application of the *second* rule for the linear variation of constants. In what follows, the *first* rule will be used in order to extend (iii_0) in a direction assigned by the applications. To this end, it will be convenient to list a few formal facts.

(I) If $X=X(t)$ is differentiable and of non-vanishing determinant at a given t, then X^{-1} is differentiable, and has the derivative $-X^{-1}X'X^{-1}$, at that t.

In fact, since the differentiability of X means the differentiability of all elements of X, it implies the differentiability of the polynomials representing $\det X$ and its minors. This proves the first assertion of (I); whence the second can be concluded by differentiating the product XX^{-1}, which is $E=\mathrm{Const.}$

(II) If $A(t)$ is continuous on a t-interval, and if $X(t)$ is a fundamental matrix of $x'=A(t)x$, then the inverse of $X^*(t)$ (which, by (14), exists) is a fundamental matrix of $x'=-A^*(t)x$.

The latter is called the adjoint of $x'=A(t)x$, the asterisk being the symbol of Hermitian transposition. In particular, $x'=A(t)x$ is self-adjoint when iA is Hermitian, that is, when $A+A^*=0$ (for every t).

The assertion of (II) is that the derivative of $(X^*)^{-1}$ is $-A^*(X^*)^{-1}$, if $X'=AX$ (and $\det X\neq 0$). But this follows from the rules

$$(X^{-1})'=-X^{-1}X'X^{-1},\quad (X^*)'=(X')^*,\quad (X^{-1})^*=(X^*)^{-1},$$

since $(AX)^*=X^*A^*$.

If $A(t)$ is continuous (on an open or infinite t-interval), and if $x'=A(t)x$ is self-adjoint, then every solution vector $x(t)$ is a bounded function of t (even if $A(t)$ is not).

This remark, no matter how trivial, contains about all that can be assured concerning "stability" in general (except when $x'=A(t)x$ can be integrated by known functions, for instance). The sufficiency of $A(t)=A^*(t)$ follows from the fact that $|x(t)|^2$ then is a first integral, since the (Hermitian) scalar product of two solutions is independent of t. This is seen by differentiation. But the true source is the general fact expressed by (II). The latter implies that, in the self-adjoint case, $(X^*)^{-1}$ is a fundamental matrix of $x'=A(t)x$, if X is. Hence, by the principle of superposition, $X=(X^*)^{-1}C$,

where C is independent of t. In particular, $X = X(t)$ is unitary (for every t) if, without loss of generality, $C = E$.

If $A = -A^*$, then every diagonal element of A, and therefore $\operatorname{tr} A$, is purely imaginary. This and the remark italicized above imply the parenthetical criterion asserted after (47) in the following theorem:

(iv_0) *Suppose that every solution $x(t)$ of $x' = A(t)x$, where $A(t)$ is continuous on an half-line, is bounded as $t \to \infty$, and that*

$$\liminf_{t\to\infty} \Re \int_{t_0}^{t} \operatorname{tr} A(s)\,ds > -\infty \tag{47}$$

(both of these assumptions are satisfied in the self-adjoint case, (47) being implied by

$$\Re \operatorname{tr} A(t) \geqq 0 \tag{47 bis}$$

as $t \to \infty$). Let $B(t)$ be any continuous coefficient matrix which is so "close" to $A(t)$ that

$$\int^{\infty} | B(t) - A(t) |\, dt < \infty. \tag{48}$$

Then, corresponding to every solution $x(t)$ of $x' = A(t)x$, there exists a solution $y(t)$ of $y' = B(t)y$ satisfying

$$y(t) - x(t) \to 0 \quad \text{as} \quad t \to \infty.$$

Moreover, this $y(t)$ is uniquely determined by $x(t)$, every solution of $y' = B(t)y$ corresponds to a solution of $x' = A(t)x$, and the correspondence is continuous in terms of the respective initial data $x(t_0)$, $y(t_0)$.

Clearly, the existence of such a correspondence will be proved if it is shown that, if $X(t)$, $Y(t)$ are fundamental matrices of $x' = A(t)x$, $y' = B(t)y$ respectively, the matrix $U(t) = U_{XY}(t)$ defined by (18) tends to a limit, $U(\infty)$, of non-vanishing determinant. In fact, the last formula line and what follows it are then implied by the principle of superposition: A matrix is a fundamental matrix of $z' = G(t)z$ if and only if it is of the form $Z(t)C$, where $Z(t)$ is an arbitrarily fixed fundamental matrix and C a unique (but unrestricted) constant matrix of non-vanishing determinant.

If A, X, $t \to +0$ in (iii_0) are replaced by K, U, $t \to \infty$ respectively (cf. the mapping (46) and the remarks following this mapping), it is seen from (iii_0) and (iii_0 bis) that the existence of a limit $U(\infty)$ of non-vanishing determinant will be proved if it is shown that

$$\int^{\infty} |K(t)|\, dt < \infty$$

is satisfied by the coefficient matrix of $u' = K(t)u$, where $U = (u_1, \cdots, u_n)$. But (13), (17) and the definition, (18), of U imply (19) and (20). Hence, it is sufficient to show that what is required by the last formula line is satisfied when K is the matrix (20).

In view of (48), this will be proved if it is ascertained that both $X(t)$ and $X^{-1}(t)$ are $O(1)$ as $t \to \infty$. But $X(t) = O(1)$ is just the assumption made before (47). This assumption also implies that every minor of $\det X(t)$ is $O(1)$. Hence, $X^{-1}(t) = O(1)$ can fail to hold only if the denominator, represented by $\det X(t)$, comes arbitrarily close to 0 as $t \to \infty$. And (15) shows that (47) excludes this possibility.

It is clear from this proof of (iv_0) that (iv_0) can be refined to a theorem, say (iv_m), which relates to (iv_0) in the same way as (iii_m) to (iii_0). The situation is made clear enough by the remark that the particular case $A = 0$ of (iv_0) is equivalent to (iii_0).

8. None of the above results contains the classical theorem, mentioned in the introduction. But if the theorem is observed to be equivalent to its wording (i), in which nothing refers to the theory of elementary divisors, there arises the question whether the classical theorem is or is not so general as to be independent of the existence of a Jordan-Weierstrass theory for the underlying space. It turns out that the answer is in the affirmative.

Needless to say, the resulting possibility of generalizing (i) in an abstract direction must be interpreted as just a symptom of the fact that neither of the classical types of proof, sketched in the introduction, takes into account the true *formal-algebraic foundations of the problem* (which, in reality, rest on the properties of the bracket operator). In other words, it will suffice to give a *direct* proof for (i) itself, since the possibility of generalizations will then become trivial. Such a proof of (i) will be seen to be contained in the following fact:

Liouvillian Lemma. *In case of n-rowed matrices, the (linear) equation*

(49) $$X - [X] = C, \quad \text{where} \quad [X] = RX - XR,$$

has a (unique) solution $X = X^R(C)$ satisfying

(50) $$|X| \leqq |C|/(1 - 2n|R|), \quad \text{if} \quad 2n|R| < 1.$$

Here R is fixed (in accordance with the hypothesis of (50)), C is arbitrary,

and the sign of absolute value is that defined before (1). A straightforward proof results by using, in the fashion of the theory of integral equations and infinite bounded matrices (Liouville, Neumann, Schwarz, Hilb), a "resolvent" series, as follows:

Let a scalar parameter, s, be inserted in front of the bracket. Then (49) becomes the case $s=1$ of

$$X - s[X] = C, \quad \text{where} \quad [X] = XR - RX.$$

If one tries to solve this equation by a power series

$$X = \sum_{m=0}^{\infty} C_m s^m,$$

where the (unknown) matrices C_m are functions of R and C, the comparison of equal powers of s clearly leads to the recursion formula $C_m - [C_{m-1}] = 0$, where $m = 1, 2, \cdots$ and $C_0 = C$. Hence, $C_m = [C]_m$, if the subscript of the bracket denotes m-fold application of (2) to $X = C$ (when R is fixed). Accordingly, if $(m)_k$ denotes the binomial coefficient,

$$C_m = \sum_{k=0}^{m} (-1)^{m-k} (m)_k R^k C R^{m-k}.$$

Hence, from (1),

$$| C_m | \leqq n^m \sum_{k=0}^{m} (m)_k \, | R |^k \, | C | \, | R |^{m-k},$$

which means that

$$| C_m | \leqq n^m \, | C | \, | R |^m \sum_{k=0}^{m} (m_k) = | C | \, | 2nR |^m.$$

In view of this inequality, the power series which has been tried for X is majorized by

$$| C | \sum_{m=0}^{\infty} | 2nsR |^m$$

and will, therefore, supply a solution, X, if this geometric progression is convergent. In the case of (49), where $s=1$, this requires the inequality $2n \, | R | < 1$ which, when satisfied, leads to

$$| X | \leqq | C | \sum_{m=0}^{\infty} (2n \, | R |)^m,$$

as claimed in (50).

The proof of (i) now becomes of such a trivial nature that the complicated analytical result of the local Fuchsian theory appears as a mere restate-

ment of the formal-algebraical fact expressed by the primitive lemma, just proved, and of the other "linear" properties of the bracket (2).

9. The assumption of (i) is that, in $W' = F(z)W$,

$$(51) \qquad F(z) - z^{-1}R = \sum_{m=0}^{\infty} A_m z^m,$$

where R and A_m are constant matrices and the power series (51) converges near $z = 0$. The assertion is that there exists a solution matrix which can be factorized into contributions, $P(z)$ and z^R, of the secondary part, (51), and of the principal part, $z^{-1}R$, of the coefficient matrix $F(z)$, in such a way that $P(z)$ in

$$(52) \qquad W(z) = P(z)z^R$$

becomes, as (51), a regular power series,

$$(53) \qquad P(z) = \sum_{m=0}^{\infty} B_m z^m,$$

which converges near $z = 0$ and has a determinant which does not vanish identically.

The latter condition requires, of course, *the existence of a (unique) non-negative integer* $l = l(R)$ such that the power series (53) becomes

$$(53 \text{ bis}) \qquad P(z) = z^l(B_l + zB_{l+1} + \cdots), \quad \text{where} \quad B_l \neq 0.$$

After this l, *the index* of the first non-vanishing B_m, has been determined, all the remaining coefficients B_m of (53) will follow *uniquely*. Finally, it will be shown that the resulting power series (53) is convergent near $z = 0$. In terms of the notation defined before (1), this means that

$$(54) \qquad |B_m| < b^m$$

holds for every m and for some $b > 0$. Correspondingly, the assumption made as to $F(z)$ is that, in (51),

$$(55) \qquad |A_m| < a^m,$$

where $a > 0$ is independent of m.

First, if (52) and (51) are substituted into $W' = F(z)W$, there results for $P(z)$ the differential equation

$$P'(z) + z^{-1}P(z)R = (z^{-1}R + \sum_{m=0}^{\infty} A_m z^m)P(z).$$

If a power series (53) is tried for $P(z)$, this differential equation can be written in the form

$$\sum_{m=1}^{\infty} mB_m z^{m-1} = z^{-1} \sum_{m=0}^{\infty} [B_m] z^m + \sum_{m=0}^{\infty} A_m z^m \sum_{k=0}^{\infty} B_k z^k,$$

where $[B_m] = RB_m - B_m R$. Hence, the comparison of the coefficients of z^{-1} and z^m, where $m = 0, 1, \cdots$, gives

$$(56) \qquad [B_0] = 0$$

and

$$(57) \qquad (m+1) B_{m+1} = [B_{m+1}] + C_m$$

respectively, where C_m is an abbreviation for the matrix which is supplied by Cauchy's multiplication of the last two power series, that is,

$$(58) \qquad C_m = \sum_{k=0}^{m} A_{m-k} B_k.$$

In order to see that this infinite sequence of conditions for the unknowns $B_0, B_1, \cdots$ can be satisfied, let the arbitrary n-rowed matrix, X, which occurs in the definition, (2), of $[X]$ be thought of as a vector with n^2 components. Then $X \to [X]$ is a linear transformation representable by an n^2-rowed matrix (which is determined by R, a fixed n-rowed matrix). Consider the characteristic numbers of this n^2-rowed matrix. If one of them happens to be the reciprocal value of a positive integer (that is, if there exists a non-negative integer, m, for which

$$X - s[X] = 0, \quad \text{where} \quad s = (m+1)^{-1},$$

has a solution $X \neq 0$), let l denote the greatest of these positive integers $(= m+1)$. Otherwise put $l = 0$. Thus $l = l(R)$ is a unique non-negative integer which, when it is positive, is characterised by the following pair of properties:

$$(59) \qquad lX - [X] = 0$$

has a solution $X \neq 0$, and, if $m + 1 > l$,

$$(60) \qquad (m+1)X - [X] = C$$

has a solution X for *every* C. And this pair of properties is characteristic of $l = l(R)$ in the case $l = 0$ also. In fact, (59) then becomes $-[X] = 0$ and has therefore the solution $X = E \neq 0$, the bracket (2) being the zero matrix

when X is the unit matrix. [It will be observed that, if $l > 0$, this definition of $l \doteq l(R)$ fails to apply in the case of infinite bounded matrices, first of all because the non-existence of a unique bounded reciprocal matrix fails to remain equivalent to the existence of a non-trivial solution of the homogeneous equations.] The coefficients B_m of (53) can now be determined as follows:

Whether $l = 0$ or $l > 0$, let B_l be a solution $X \neq 0$ of (59). In the first case, this implies that (56) is satisfied. In the second case, that is, if there exists at least one B_m preceding B_l, let $B_0 = 0, \cdots, B_{l-1} = 0$. Then (56) remains fulfilled, since the matrix (2) is 0 when $X = 0$, and (58) becomes satisfied for every non-negative $m < l$, the corresponding matrices C_m being $C_0 = 0, \cdots, C_{l-1} = 0$. Consequently, (57) is satisfied for every non-negative $m < l - 1$ (provided that there exists such an m, i. e., that $l > 1$). But (57) is satisfied for $m = l - 1$ also (whether $l > 1$ or $l = 1$), since, in view of $C_{l-1} = 0$, the condition assigned for $B_{m+1} = B_l$ by (57) is identical with the assumption that $X = B_l$ is a solution of (59).

Accordingly, whether $l = 0$ or $l > 0$, the $l + 1$ matrices $B_0, \cdots, B_l$ are defined so as to satisfy the initial condition (56) and those of the equations (57) and (58) which belong to any non-negative $m < l$ (provided that there exists such an m, i. e., that $l > 0$). Now let $m = l$ (whether $l > 0$ or $l = 0$). Then (58) defines C_l as a function of $B_0, \cdots, B_l$. Hence, the case $m = l$ of (57) becomes a condition for B_{l+1}. But this condition has a unique solution, since (60) has a unique solution $X = X^C$ whenever $m + 1 > l$. And it is now clear that the possibility of a complete induction, leading from a given B_m to the corresponding B_{m+1}, is never arrested.

This proves the existence of a (formal) power series (53 bis). What remains to be shown is that this power series has a non-vanishing radius of convergence.

To this end, let (57) be multiplied by $(m + 1)^{-1}$, and let $m + 1$ be replaced by m. Then it is seen from (2) that the resulting representation of (57) can be written in the form (49), by choosing

$$X,\ C,\ R \quad \text{to be} \quad B_m,\ m^{-1}C_{m-1},\ m^{-1}R$$

respectively. It follows therefore from the Liouvillian lemma, that B_m (exists, is unique and) is subject to the inequality

$$|B_m| \leqq |C_{m-1}|/(m^2 - 2n|R|),$$

if the proviso of (50) is satisfied, that is, if the denominator is positive. In the present case, this proviso is satisfied whenever the first term, m^2, of the denominator is large enough, since the second term, $-2n|R|$, depends only

on the dimension number and the residue, which are fixed. Consequently, there exists a positive constant, say d, satisfying

$$|B_m| \leqq d\,|C_{m-1}|/m^2$$

from a certain $m = m_0$ onward.

In particular, if d is chosen large enough, $|B_m| \leqq d\,|C_{m-1}|$ holds from $m = 0$ onward (provided that C_{-1} is declared to denote 1, for instance). Hence, it is seen from (58) and (1) that

$$|C_m| \leqq nd \sum_{k=0}^{m} |A_{m-k}|\;|C_{k-1}|.$$

Consequently, if c_m denotes the greatest of the values $|C_{k-1}|$, where $k \leqq m$, then the last two formula lines imply that $|B_m| = O(c_m)$ and

$$c_{m+1} \leqq ndc_m \sum_{k=0}^{m} a^{m-k},$$

by (55). Hence,

$$c_{m+1} = ndc_m O(\alpha^m), \quad \text{where} \quad \alpha = \max(2, a),$$

and so, by induction,

$$c_m = O(\beta^m), \quad \text{where} \quad \beta = nd\alpha.$$

It follows therefore from $|B_m| = O(c_m)$ that (54) is satisfied by some $b > 0$, which proves that (53) has a circle of convergence.

10. For a definitive nature of the formal algebra in the above treatment of a Fuchsian point, a test case quite different from the resulting possibility of extending the Fuchsian theory to generalized linear spaces will now be considered. In fact, the Fuchsian character of the singular point, that is, the assumption of a vanishing *rank* (Poincaré), will now be transferred to the non-degenerate case of a positive rank, leading to *normal series* (Thomé).

In this case of (in general) divergent expansions, the standard formal complications arising from a multiple invariant factor, and (even if all these factors are simple) from a characteristic exponent which is multiple (mod 1), are not usually treated in detail. But it turns out that, in the proof for the existence of the full number of formal expansions, the classical method of counting the constants, that is, the machinery of the theory of elementary divisors, achieves, again, more harm than good. In fact, a more direct procedure is able to lead to explicit results corresponding to the factor $z^R = \exp(R \log z)$ in (52).

The reasons for the above wording, (i), of the theorem dealing with the case of rank 0, and for the possibility of a straightforward verification of this wording, were two-fold: The first factor in (52), that is, the perturbation due to (51) was automatically prescribed by the "second method of the variations of constants," and the explicit factor, z^R, in (52) was a fundamental matrix of the undisturbed system. In view of (51), the latter belongs to the coefficient matrix $F(z) = z^{-1}R$, that is, to Cauchy's elementary system, $zW' = RW$, where $R = \text{Const.}$ Precisely this system and this reasoning are the formal foundations of the non-local theory of Schlesinger, quoted above.

Correspondingly, the prototype of a singularity of arbitrary rank, say μ, is the singularity of the system $z^{1+\mu}W' = RW$ (at $z = 0$), where $R = \text{Const.}$ But a differentiation verifies that a solution matrix, $W = W(z)$, of this prototype is simply

$$(61) \qquad R_\mu(z) = \exp(-\mu^{-1}z^{-\mu}R)$$

(if $\mu \neq 0$; if $\mu = 0$, then $-\mu^{-1}z^{-\mu}$ must be interpreted as its limit when $\mu \to 0$, that is, as $\log z$; so that $R_0(z)$ becomes $\exp(R \log z) = z^R$). Consequently, if the trivial coefficient matrix of the prototype is disturbed somewhat, the "second method of the variation of constants" assigns a fundamental matrix of the form $P(z)R_\mu(z)$, where the matrix P (which *must* be written in front of R_μ) represents the perturbation.

It turns out that this plan, leading to *an explicit representation*, (61), *of the principal terms in a complete set of linearly independent "normal" (or "anormal") series*, can be carried out without any difficulty. The reason is precisely the avoidance of characteristic numbers and elementary divisors, which only disguise the simple exponential factor, (61). In other words, the leading terms of the expansions, the terms responsible for the formal difficulties of the usual treatment, can be split off *en bloc*, since they happen to be identical with the exact solution, (61), of the elementary prototype. This is the content of the following formal extension of (i), where $\mu = 0$, to the case of an arbitrary rank μ:

(i*) *If the n-rowed matrix F is an analytic function having a pole at the point $z = 0$ of the z-plane, then $w' = F(z)w$ has a formal solution matrix of the form $P(z)R_\mu(z)$, where the matrix $P(z)$ is a power series* (53), *with a (unique) non-negative $l = l(R)$ in* (53 bis), *and the second factor is the exponential matrix* (61) *in which $\mu + 1$ denotes the order of the pole, and* $R = \text{Const.}$ *the coefficient of the leading term, of $F(z)$ at $z = 0$ (that is, $z^\mu F(z)$ has a simple pole, of residue R, at $z = 0$, the subscript of* (61) *being identical with Poincaré's non-negative index of rank).*

In other words, the assumption is that

$$F(z) - z^{-1-\mu}R = H(z),$$

where $H(z)$ has at $z = 0$ a pole of order $\mu (\geqq 0)$, at most. And the assertion is that $W' = F(z)W$ can formally be satisfied by

$$W(z) = P(z)R_\mu(z),$$

where $P(z)$ is some power series of the form (53), (53 bis).

First, the substitutions defined by the last two formula lines transform $W' = FW$ into

$$P'R_\mu + PR_\mu' = z^{-1-\mu}RPR_\mu + HPR_\mu.$$

Since, by (61),

$$R_\mu' = z^{-1-\mu}RR_\mu,$$

it follows that the differential equation to be satisfied by $P = P(z)$ can be written in the form

$$P' = z^{-1-\mu}[P] + H(z)P,$$

if use is made of the abbreviation (2).

If $\mu = 0$, the assertion of (i*) is contained in that of (i). Since, as will be clear from the proof, there is no difference between the treatment of $\mu = 1$ and that of an arbitrary $\mu > 0$, let the formulae be curtailed by choosing $\mu = 1$. Then, the pole of $H(z)$ being of order μ (at most),

$$P' = z^{-2}[P] + z^{-1}SP + G(z)P,$$

where S denotes the residue of $H(z)$ at $z = 0$ and $G(z)$ is a regular power series,

$$G(z) = \sum_{m=0}^{\infty} A_m z^m.$$

Thus, from (53) and (2),

$$\sum_{m=0}^{\infty} mB_m z^{m-1} = z^{-2} \sum_{m=0}^{\infty} [B_m] z^m + z^{-1}S \sum_{m=0}^{\infty} B_m z^m + \sum_{m=0}^{\infty} A_m z^m \sum_{k=0}^{\infty} B_k z^k.$$

If this is written in the form

$$z^{-1} \sum_{m=1}^{\infty} mB_m z^m = z^{-1} \sum_{m=0}^{\infty} [B_m] z^m + \sum_{m=0}^{\infty} SB_m z^m + z \sum_{m=0}^{\infty} (\sum_{k=0}^{\infty} A_{m-k} B_k) z^m,$$

then, since $S = \text{Const.}$, the comparison of the coefficients of z^{-1}, z^0 and $z^1, z^2, \cdots$ is seen to lead to

$$0 = [B_0], \qquad B_1 = [B_1] + SB_0$$

and

$$mB_m = [B_m] + SB_{m-1} + \sum_{k=0}^{m-2} A_{m-2-k}B_k$$

respectively, where $m = 2, 3, \cdots$. But the first of these three conditions is precisely (56), whereas the second and the third can be united, and appear in the form (57), where $m+1$ commences with $m+1=1$ and, in view of the last two formula lines, C_m must be defined as follows:

(62) $$C_0 = 0, \qquad C_m = SB_m + \sum_{k=0}^{m-1} A_{m-1-k}B_k \quad \text{if} \quad m > 0.$$

Accordingly, the full system of conditions is just the same as above, since it consists of (56) and (57), where $m = 0, 1, \cdots$. However, (58) must now be replaced by (62). But this shift in the structure of C_m prevents a repetition of the proof of convergence by means of a dominating geometrical progression, as given above for the case, $\mu = 0$, of (i). And, as is well known, the radius of convergence of the resulting power series (53) *is* 0 even in the simplest examples of rank $\mu = 1 \neq 0$.

Correspondingly, (i*) does not contain any statement concerning the existence of actual solutions (functions) to which the formal power series "belong." However, it could be proved that *the formal power series,* (53 bis), *is summable in Borel's sense* (if $\mu = 1$, and in the sense of Le Roy's generalized Laplace-Borel transforms, if μ is arbitrary).

11. In his fundamental paper referred to above (and, at least between the lines, in his earlier investigations which he mentions but does not presuppose there), Schlesinger is led to the quadratic system of $k+1$ matrix differential equations

(63) $$\begin{aligned} X_0' &= \sum_{m=1}^{k} (t - a_m)^{-1}[X_0, X_m] \qquad &([A, B] = AB - BA) \\ X_m' &= (t - a_m)^{-1}[X_m, X_0]; \qquad &m = 1, \cdots, k, \end{aligned}$$

where $a_1, \cdots, a_k$ are (distinct) points of the complex t-plane and $X_0, \cdots, X_m$ are n-rowed matrices. Hence, in terms of scalars, x_i,

(64) $$x_i' = L_i(x_1, \cdots, x_p; t) + Q_i(x_1, \cdots, x_p; t),$$

where the order, p, of the system is $(k+1)n^2$, and L_i and Q_i denote linear and quadratic forms respectively (actually, L_i vanishes identically). However, it is clear that (63) has the linear integral

$$\sum_{j=0}^{k} X_j(t) = \text{const.}, \tag{63 bis}$$

by means of which it is reducible to a system of the form (64) and of order $p = kn^2$.

By connecting (63) with his *schlechthin* canonical linear systems of the Fuchsian type, Schlesinger recognizes in the quadratic system (63) a class of non-linear differential equations (substantially of order mn^2) which, precisely because of its connection with the Riemann-Fuchs problem, has only solutions with fixed "critical" singularities. In other words, all movable singularities of any of the functions $x_i(t)$ (that is, those of their singular points, $t = t^0$, which depend on integration constants) are poles. And the fact is that, no matter how transparent the function-theoretical situation may be, the formulation (64) of (63) exhausts, in case of the lowest values of p, all types contained in the work of Painlevé and his pupils, and leads, with increasing p, to an infinity of new "Painlevé transcendents."

It seems to be worth observing (if it has not been observed before), that there is a class of systems of the same type as (63) but depending on purely formal considerations, rather than on function-theoretical arguments. The systems in question are again of the form (64), result again from linear matrix differential equations, and all the critical singularities are again fixed, but this time for an *explicit* reason: The solutions $X(t)$ depend on the reciprocal matrices $U^{-1} = U^{-1}(t)$ of solutions $U = U(t)$ of a linear differential equation of the second order for the matrix U (cf. (70) and (71) below), and so the only movable singular points, $t = t^0$, of the elements, $x_i(t)$, of the matrix $X(t)$ are those at which the determinant of a particular $U(t)$ happens to vanish; a situation corresponding to that dealt within the theory of conjugate points.

12. The nature of the class in question being of a formal origin, it is unnecessary to assume analyticity. In fact, all that is needed is a transcription of Riccati's quadratic equation,

$$x' = a(t) + b(t)x + c(t)x^2, \tag{65}$$

to the case of matrices. The sole trick is that the quadratic term, $c(t)x^2$, in (65) must then be written as $xc(t)x$:

$$X' = A(t) + B(t)X + XC(t)X. \tag{66}$$

If the matrices are n-rowed, (66) is a system (64) of order $p = n^2$.

In order to simplify the situation, suppose first that the third coefficient of (66) is the unit matrix:

$$X' = A(t) + B(t)X + X^2. \tag{67}$$

As to $A(t)$ and $B(t)$, it is sufficient to assume mere continuity on a t-interval.

Every pair of initial vectors, say $u(t_0)$ and $u'(t_0)$, determines a solution vector $u = u(t)$ of

$$u'' - B(t)u' + A(t)u = 0. \tag{68}$$

If $u = u_1(t), \cdots, u = u_n(t)$ are n solution vectors of (68), the matrix

$$U(t) = (u_1, \cdots, u_n) \tag{69}$$

is a solution of

$$U'' - B(t)U' + A(t)U = 0, \tag{70}$$

and conversely. However, even if the columns of (69) are linearly independent, the n-rowed matrix (69) cannot be a fundamental matrix of (68), since (68), being of order $2n$, has $2n$, instead of just n, linearly independent solutions $u = u(t)$. In particular, $\det U(t)$ can have an isolated zero, say $t = t^0$, within the t-interval of continuity (or, for that matter, regular-analyticity) of the coefficient functions $A(t)$, $B(t)$.

It is precisely the possibility of such zeros $t = t^0$ that leads to "movable singularities" of $X(t)$. All the other "singular" points of $X(t)$ are "fixed," namely, such as to be "singularities" of the coefficient functions $A(t)$, $B(t)$ themselves. For, on the one hand, (70) is linear and, on the other hand, the connection between (70) and (67) is simply as follows:

$$X = -U'U^{-1}. \tag{71}$$

Here $U = U(t)$ denotes any set (69) of n linearly independent solutions of (68). Hence, $\det U(t)$ does not vanish *identically*. Consider a t-interval on which $\det U(t)$ has no zero. On such an interval, (71) defines a function $X = X(t)$. The latter has a continuous first derivative, since $U = U(t)$, being a solution of (70), has a continuous second derivative. But, since $(U^{-1})' = -U^{-1}U'U^{-1}$ (cf. p. 199 above), the derivative of (71) is

$$X' = -U''U^{-1} + U'U^{-1}U'U^{-1}.$$

Hence, direct substitutions show that (67) becomes an identity in t by virtue of (71) and (70).

In the more general case, (66), the assumption of mere continuity remains sufficient for $A(t)$ and $B(t)$, but $C(t)$ must be assumed to have a continuous first derivative and a non-vanishing determinant. In fact, (71) must now be replaced by

$$X = -C^{-1}U'U^{-1}, \tag{72}$$

and (70) by

$$C(t)U'' - {}^*B(T)U' + A(t)U = 0, \tag{73}$$

where ${}^*B$ is an abbreviation for

$$ {}^*B = BC^{-1} + C^{-1}C'C^{-1}. \tag{74}$$

The verification is the same as before, the derivative of (72) being

$$X' = DU'U^{-1} - CU''U^{-1} + CU'U^{-1}U,$$

where D denotes the second term, $C^{-1}C'C^{-1}$, of (74).

It will be observed that, in view of (69), the system (70) is equivalent to

$$C(t)u'' - {}^*B(t)u' + A(t)u = 0,$$

a linear system of order $2n$, whereas (66) is a quadratic system of order n^2. Thus, the reduction effected by Riccati's substitution, (72), in the classical case, $n = 1$, is quite accidental, since $2n$ becomes equal to n^2 when $n = 2$, and is less than n^2 from $n = 3$ onward. It follows that, in contrast to the classical case, the connection between (66) and (73) does not play the rôle of "reducing the order," a rôle somewhat insignificant since about Riemann's time, but rather the rôle of a function-theoretical reduction, leading, via a linear system, to a non-linear system with fixed critical singularities.

The exceptional standing of the scalar case in question concerning a reduction of

$$y'' + P(t)y' + Q(t)y = 0 \tag{75}$$

can also be seen from what results if a "Tschirnhaus transformation" is tried for (75), where both n-rowed matrices, $P(t)$ and $Q(t)$, are given as continuous functions on a t-interval. In fact, if $n = 1$, just the quadrature of $P(t)$ is needed in order to transform (75) into an equation, say

$$z'' + R(t)z = 0, \tag{76}$$

in which the first derivative does not occur. But if $n > 1$, what is needed to this end is the general solution of

$$u' = -\tfrac{1}{2}P(t)u, \tag{77}$$

that is, of a system of order n (so that (76) and (77) together represent a system of order $3n$, whereas (75) itself is of order $2n$). That the general solution of (77), that is, a fundamental matrix (69), actually leads to a reduction of (75) to (76), can be verified by what amounts to an application

of the variation of constants which, however, fails under the mere assumption of continuity for the matrices $P(t)$, $Q(t)$.

Suppose therefore that $P(t)$ has a continuous first derivative. Then every fundamental matrix, $U(t)$, of (77) has a continuous second derivative. In order to "vary the constants," put

$$Y = UZ, \tag{78}$$

where $Y = Y(t)$ denotes a matrix formed by n linearly independent solution vectors of (75). Then $Z = Z(t)$ supplies a corresponding solution of (76), where the coefficient matrix is the continuous function $R = R(t)$ defined by

$$R = -\tfrac{1}{2}U^{-1}(P' - \tfrac{1}{2}P^2 - 2Q)U \tag{79}$$

(U^{-1} exists since $U = U(t)$ is a fundamental matrix).

In fact, from (78),

$$Y' = U'Z + UZ', \qquad Y'' = U''Z + 2U'Z' + UZ''.$$

If this pair of relations and (78) are substituted into what results when y is replaced by Y in (75), then, since the assumption (77) is equivalent to $2U' + PU = 0$, it follows that

$$UZ'' + (U'' + PU' + QU)Z = 0.$$

Since

$$U' = -\tfrac{1}{2}PU, \quad \text{hence} \quad U'' = -\tfrac{1}{2}(P' - \tfrac{1}{2}P^2)U,$$

this can be written in the form $Z'' + RZ = 0$ or (76), if R is defined by (79).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

The report of E. Hilb, "Lineare Differentialgleichungen im komplexen Gebiet," *Encyklopädie der mathematischen Wissenschaften*, vol. II_2, pp. 471-562, more particularly pp. 473-494, gives a full account (up to 1915) of the extended literature of the local problems referred to above. As to the classical beginnings of the theory, cf. also the comments of O. Haupt at the end of his edition of F. Klein's *Vorlesungen über die hypergeometrische Funktion*, Berlin, 1933.

FIXED POINT THEOREMS FOR MULTI-VALUED TRANSFORMATIONS.*

By SAMUEL EILENBERG and DEANE MONTGOMERY.

1. Introduction. Recently there have been several extensions of known fixed point theorems in which the transformation T takes each point of a compact metric space M into a closed subset of M. For such a transformation a point x is said to be a fixed point if x is in $T(x)$. These extensions first occurred in von Neumann's work on the theory of games (see [8] where earlier references are also given). Kakutani [3] proved a theorem which we shall formulate below, and Wallace [9] also has a theorem in this direction.

Our purpose in the present paper is to present a general fixed point theorem which on the one hand includes a very general form of the famous fixed point formula of Lefschetz [4], [5], [6] and on the other hand implies the fixed point theorems of Kakutani and Wallace.

We rely heavily on a theorem proved by Vietoris [7] and with this as a tool the proof for the case at hand resembles the proof given by Lefschetz.

2. Definitions and theorems. Let M and N be metric compact spaces and let $T: M \rightarrow N$ be a multi-valued function, i. e., a correspondence which to each $x \in M$ assigns one or more points of N. For every $x \in M$, $T(x)$ will denote the set of all "images" of x. The function T is continuous provided $x_n \rightarrow x$, $y_n \rightarrow y$ and $y_n \in T(x_n)$ imply $y \in T(x)$. The *graph* Γ of T is the subset of the cartesian product $M \times N$ consisting of points (x, y) with $y \in T(x)$. The continuity of T is equivalent with the condition that Γ is closed. The continuity of T implies that the sets $T(x)$ are closed. If we regard T as a point-to-set function then the continuity of the multi-valued function T is equivalent with the upper semi-continuity of the set-valued function.

Unless otherwise stated, we shall use Vietoris cycles and homologies over a field F of coefficients.

Definition. A compact metric space X is said to be *acyclic* provided 1°) X is non-vacuous, 2°) the homology groups $H_q(x)$ vanish for $q > 0$, 3°) the reduced 0-th homology group $\tilde{H}_0(x)$ vanishes.

* Received January 4, 1946.

The reduced 0-th homology group is obtained by considering only cycles in which the sum of coefficients is zero.

THEOREM 1. *Let M be an acyclic absolute neighborhood retract and $T: M \to M$ a continuous multi-valued function such that for every $x \in M$ the set $T(x)$ is acyclic. Then T has a fixed point.*

Since every compact convex set in a Euclidean space is an absolute neighborhood retract and is acyclic, Theorem 1 yields:

KAKUTANI'S THEOREM. *Let M be a compact convex subset of Euclidean n-space and let $T: M \to M$ be a continuous multi-valued function such that for every $x \in M$ the set $T(x)$ is convex. Then T has a fixed point.*

Wallace's theorem (in the metric case) follows from Theorem 1 by assuming that M is a tree. We also note that the case when M is an acyclic (finite) polyhedron is included in Theorem 1 since polyhedra are absolute neighborhood retracts.

3. Reformulation of the problem. We first formulate Theorem 1 in a somewhat more general fashion.

THEOREM 2. *Let M be an acyclic absolute neighborhood retract, N a compact metric space, $r: N \to M$ a continuous single valued mapping and $T: M \to N$ a multi-valued continuous mapping such that all the sets $T(x)$ are acyclic for $x \in M$. Then the combined (multi-valued) mapping $rT: M \to M$ has a fixed point.*

Taking $N = M$ and $r(x) = x$ for all $x \in M$ yields Theorem 1.

Definition. A continuous mapping $f: N \to M$, where N and M are compact metric spaces will be said to have property (V) provided for every $x \in M$ the antecedent set $f^{-1}(x)$ is acyclic.

It follows that if f satisfies property (V) then it maps N *onto* M.

Using the property (V) we can replace Theorem 2 by an equivalent theorem involving only singled valued transformations.

THEOREM 3. *Let M be an acyclic absolute neighborhood retract, N a compact metric space and $r: N \to M$, $t: N \to M$ continuous mappings. If t satisfies property (V) then r and t have a coincidence, i. e., $r(x) = t(x)$ for some $x \in N$.*

We first show that Theorem 3 implies Theorem 2. Let M, N, r and T

satisfy the conditions of Theorem 2. Consider in the cartesian product $M \times N$, the subset N^* consisting of all points of the form (x, y) where $x \in M$, $y \in T(x)$. Define the mappings $r^*: N^* \to M$ and $t^*: N^* \to M$ by setting

$$r^*(x, y) = r(y), \qquad t^*(x, y) = x.$$

Since $t^{*-1}(x)$ is homeomorphic with $T(x)$, the mapping t^* satisfies property (V). Hence by Theorem 3 r^* and t^* have a coincidence; this means that there is an $x \in M$ and a $y \in T(x)$ such that $r(y) = x$. Consequently $x \in T(x)$.

Conversely suppose that Theorem 2 holds. Let M, N, r and t satisfy the conditions of Theorem 3. Define $T(x) = t^{-1}(x)$ for each $x \in M$. It is then clear that M, N, r and T satisfy the conditions of Theorem 2 and there is a point $x \in M$ such that $x \in rT(x)$. Hence the sets $r^{-1}(x)$ and $t^{-1}(x) = T(x)$ intersect. Let $x' \in r^{-1}(x) \cap t^{-1}(x)$; then $r(x') = t(x')$.

4. The Lefschetz number. Given a continuous mapping $f: N \to M$ of a compact metric space N into another such space M we shall denote by f_i the homomorphism of the homology groups, $f_i: H_i(N) \to H_i(M)$, induced by f.

The following theorem established by Vietoris [7] is of fundamental importance here.

VIETORIS' THEOREM. *If $t: N \to M$ has property* (V) *then t_i maps $H_i(N)$ isomorphically onto $H_i(M)$.*

We are now in a position to formulate the theorem that we propose to prove.

THEOREM 4. *Let M be an absolute neighborhood retract, N a compact metric space, $r: N \to M$, $t: N \to M$ continuous mappings of which t satisfies property* (V). *Consider the Lefschetz number:*

$$\Lambda(r, t) = \Sigma(-1)^i \text{ trace } (r_i t_i^{-1}).$$

If $\Lambda(r, t) \neq 0$ then r and t have a coincidence.

We first note that t_i^{-1} is defined in view of Vietoris' Theorem. Further since M is an absolute neighborhood retract $H_i(M)$ is a finite dimensional vector space (over the coefficient field) and the trace of the homomorphism

$$r_i t_i^{-1}: H_i(M) \to H_i(M)$$

is defined. Finally since $H_i(M) = 0$ for sufficiently large i the summation $\Sigma(-1)^i$ trace $(r_i t_i^{-1})$ is actually finite and $\Lambda(r, t)$ is defined.

If M is acyclic then $H_i(M)=0$ for $i>0$ and $\Lambda(r,t)=1$. Hence Theorem 4 implies Theorem 3.

In the next section we shall prove Theorem 4 assuming that M is a polytope. From this partial result we derive in Section **6** Theorem 4 in full generality.

Before proceeding we point out that Theorem 4 implies Theorem 5 which is an extension of the Lefschetz fixed point formula. Let M be an absolute neighborhood retract and let $T: M \to M$ be a continuous multi-valued function such that $T(x)$ is acyclic. In the product of M and M let N be the graph of T. Define the mappings $r: N \to M$ and $t: N \to M$ as follows:

$$r(x,y)=y, \quad t(x,y)=x.$$

Then t satisfies property (V) and we may form the Lefschetz number

$$\Lambda(T)=\Lambda(r,t)=\Sigma(-1)^i \text{ trace } (r_i t_i^{-1}).$$

Theorem 4 implies

THEOREM 5. *Let M be an absolute neighborhood retract and let $T: M \to M$ be a continuous multi-valued function such that $T(x)$ is acyclic for each $x \in M$. If $\Lambda(T) \neq 0$ then T has a fixed point.*

5. Proof of Theorem 4 for polyhedra. Assume that M is a finite polyhedron given in a simplicial decomposition, $\mathfrak{M}$. For each $x \in M$ denote by $v(x)$ a vertex of the lowest dimensional simplex containing x. We may select $\epsilon > 0$ such that

(1) for every ϵ-chain c in M, $v(c)$ is a chain on $\mathfrak{M}$.

Assume now that r and t do not have a coincidence; we may then select the simplicial decomposition $\mathfrak{M}$ so fine that

(2) for each $x \in N$ the vertices $v(r(x))$ and $v(t(x))$ are not in the same simplex of $\mathfrak{M}$.

Since r and t are continuous we may select $\epsilon_1 > 0$ so that

(3) for every ϵ_1-chain c in N, $r(c)$ and $t(c)$ are ϵ-chains in M.

Further it follows from Vietoris' theorem that we may select ϵ_2 such that $0 < \epsilon_2 < \epsilon_1$ with the property that

(4) if c is an ϵ_2-cycle in N and $vt(c) \sim 0$ in $\mathfrak{M}$, then $c \underset{\epsilon_1}{\sim} 0$ in N.

Let s be any simplex of $\mathfrak{M}$. Since s is acyclic it follows from Vietoris'

theorem that $f^{-1}(s)$ is acyclic. Hence we may select a sequence $0 < \eta_1 < \cdots < \eta_n < \epsilon_2$ where n is the dimension of $\mathfrak{M}$ such that

(5) if c is a $(q-1)$-dimensional η_{i-1}-cycle in $t^{-1}(s)$, where s is any q-simplex of $\mathfrak{M}$, then $c \underset{\eta_i}{\sim} 0$ in $t^{-1}(s)$.

For $q = 1$ we assume in (5) that the zero-dimensional cycle c has the sum of coefficients zero.

We shall now define a chain transformation τ with the following properties.

(6) For every i-dimensional chain c on $\mathfrak{M}$, τc is an i-dimensional η_i-chain in N.

(7) $\partial\tau(c) = \tau(\partial c)$.

(8) $\tau(c_1 + c_2) = \tau(c_1) + \tau(c_2)$.

(9) $\tau(fc) = f\tau(c)$ for every f in the coefficient field F.

(10) $\tau(c) \subset t^{-1}(|c|)$, where $|c|$ is the smallest sub-complex of $\mathfrak{M}$ containing c.

(11) $vt\tau(c) = c$.

If c is a vertex in $\mathfrak{M}$ taken with multiplicity one, we define $\tau(c)$ to be any point of $t^{-1}(c)$, also taken with multiplicity one. Conditions (8) and (9) then, determine $\tau(c)$ for every 0-chain c. Suppose that the construction of $\tau(c)$ has been carried out for dimensions $< i$ and let c be an oriented i-dimensional simplex of $\mathfrak{M}$ taken with multiplicity 1. Since $\partial[\tau(\partial c)] = \tau(\partial\partial c) = 0$, it follows that $\tau(\partial c)$ is an $(i-1)$-dimensional η_{i-1}-cycle in $t^{-1}(|c|)$. Hence by (5) there is an η_i-chain τc in $t^{-1}(|c|)$ such that $\partial\tau(c) = \tau(\partial c)$. Further $vt\tau(c)$ is a chain of $\mathfrak{M}$ on $|c$ and $\partial[vt\tau(c)] = vt\tau(\partial c) = \partial c$. Hence $vt\tau(c) = c$. Using (8) and (9) the definition is extended to arbitrary i-dimensional chains on $\mathfrak{M}$.

Define

$$\rho c = v r \tau(c) \tag{12}$$

for every chain c on $\mathfrak{M}$. Since $\eta_i < \epsilon_2 < \epsilon_1$ it follows from (6), (3) and (1) that ρc is a well defined chain on $\mathfrak{M}$. By (7), (8) and (9), ρ is a chain transformation, i. e.,

$$\partial\rho(c) = \rho\partial(c), \quad \rho(c_1 + c_2) = \rho(c_1) + \rho(c_2), \quad \rho(fc) = f\rho(c).$$

Consider the homomorphisms on the groups of chains

$$\rho_i : C_i(\mathfrak{M}) \to C_i(\mathfrak{M})$$

and the induced homomorphisms on the homology groups

$$\bar{\rho}_i : H_i(\mathfrak{M}) \to H_i(\mathfrak{M}).$$

The well known argument involving the additivity of the traces implies [1], [5]

$$\Sigma(-1)^i \text{ trace } \rho_i = \Sigma(-1)^i \text{ trace } \bar{\rho}_i.$$

From (11), (12) and (2) it follows that trace $\rho_i = 0$. Hence

$$\Sigma(-1)^i \text{ trace } \bar{\rho}_i = 0. \tag{13}$$

Let z be any cycle in $\mathfrak{M}$. By Vietoris' theorem there is a convergent cycle $Z = (z_1, z_2 \cdots)$ in N such that $t(Z) \sim z$. For sufficiently large n, $\tau(z) - z_n$ is an ϵ_2-cycle in N and $vt(\tau(z) - z_n) = vt\tau(z) - vtz_n = z - vtz_n \sim 0$. Therefore by (4), $\tau(z) \underset{\epsilon_1}{\sim} z_n$ in N. Consequently, by (1), $\rho(z) = vr\tau(z) \sim vrz_n$ in $\mathfrak{M}$. This shows that $\rho(z) \sim r(Z)$ while $z \sim t(Z)$. If we recall the definition of the homomorphisms r_i and t_i, this shows that $\bar{\rho}_i = r_i t_i^{-1}$. Hence (13) implies that

$$\Sigma(-1)^i \text{ trace } (r_i t_i^{-1}) = 0$$

contrary to assumption.

6. Completion of the proof. As a preliminary we dispose of the case when M is the cartesian product of a polytope P and the Hilbert cube Q_ω. By considering the first n-coördinates of Q_ω we represent Q_ω as the cartesian product $Q_n \times R_n$ of an n-dimensional cube Q_n and a Hilbert cube R_n. Define $P_n = P \times Q_n$; then $M = P_n \times R_n$. Let $\pi_n : M \to P_n$ be the natural projection $\pi_n : (x, y) \to x$. For every $x \in P_n$ the set $\pi_n^{-1}(x)$ is homeomorphic with R_n. Hence π_n satisfies condition (V). Consider now the maps

$$r : N \to M, \qquad t : N \to M$$

and the maps

$$r_n = \pi_n r : N \to P_n, \qquad t_n = \pi_n t : N \to P_n.$$

Since both π_n and t satisfy condition (V), so does the combined map $t_n = \pi_n t$. Since by Vietoris' theorem π_{ni} maps $H_i(M)$ isomorphically onto $H_i(P_n)$ we have

$$r_{ni} t_{ni}^{-1} = \pi_{ni} r_i t_i^{-1} \pi_{ni}^{-1}.$$

Hence trace $(r_{ni} t_{ni}^{-1}) =$ trace $(r_i t_i^{-1})$ and therefore $\Lambda(r, t) = \Lambda(r_n\ t_n)$.

Assume now that $\Lambda(r, t) \neq 0$. Since P_n is a polytope and $\Lambda(r_n, t_n) \neq 0$, r_n and t_n have a coincidence. Hence $\pi_n r(x_n) = \pi_n t(x_n)$ for some $x_n \in N$. If x is a limit point of the set $\{x_n\}$, it follows that $r(x) = t(x)$.

We now pass to the general case when M is an absolute neighborhood retract. We may assume that M is a subset of a larger space D with the following properties [2]:

(1) D is the cartesian product of some polytope with a Hilbert cube,

(2) M is a retract of D

Let $\rho: D \to M$ be a map retracting M to D (i.e., $\rho(x) = x$ for $x \in M$).

Consider in the cartesian product $D \times N$ the set N_1 consisting of all points (x, y) such that $x \in M$ and $t(y) = x$. For (x, y) in N_1 define $r_1(x, y) = r(y)$ and $t_1(x, y) = t(y) = x$. The map $x \to (x, t(x))$ sets up a homeomorphism $N \to N_1$, $t \to t_1$, $r \to r_1$ and therefore

$$\Lambda(r, t) = \Lambda(r_1, t_1). \tag{3}$$

Next consider in $D \times N$ the set N_2 consisting of points (x, y) satisfying the condition $\rho(x) = t(y)$. For (x, y) in N_2 define

$$r_2(x, y) = r(y), \qquad t_2(x, y) = x.$$

Clearly $r_2: N_2 \to D$ and $t_2: N_2 \to D$. For $x \in D$ the set $t_2^{-1}(x)$ is the set of all pairs $(x, y) \in D \times N$ satisfying the condition $\rho(x) = t(y)$. Hence $t_2^{-1}(x)$ is homeomorphic with $t^{-1}(\rho(x))$, and therefore t_2 satisfies condition (V).

We shall prove that

$$\Lambda(r_2, t_2) = \Lambda(r_1, t_1). \tag{4}$$

First note that $N_1 \subset N_2$ and that $r_1 = r_2$ and $t_1 = t_2$ on N_1. Since M is a retract of D, every cycle of M that bounds in D also bounds in M. Consequently we may regard $H_i(M)$ as a subgroup of $H_i(D)$. Consider the homomorphisms

$$r_{2i}t_{2i}^{-1}: H_i(D) \to H_i(D), \qquad r_{1i}t_{1i}^{-1}: H_i(M) \to H_i(M).$$

These two homomorphisms agree on $H_i(M)$. Moreover since r_2 maps N_2 into M the image of the homomorphism $r_{2i}t_{2i}^{-1}$ is in the subgroup $H_i(M)$. Hence the theorem on the additivity of traces implies trace $(r_{2i}t_{2i}^{-1}) =$ trace $(r_{1i}t_{1i}^{-1})$ and (4) follows.

Assume now that $\Lambda(r, t) \neq 0$. From (3) and (4) it then follows that $\Lambda(r_2, t_2) \neq 0$. Because of (1) the theorem can be applied to yield a coincidence for r_2 and t_2. Hence there is a pair $x \in D$, $y \in N$ such that

$$\rho(x) = t(y), \qquad r(y) = x.$$

The second condition implies that $x \in M$, hence $\rho(x) = x$ and $r(y) = t(y)$, q. e. d.

7. A special case. In the special case when M is a euclidean n-cell Theorem 1 can be given a more direct and geometric proof leading to a slightly more general result.

We shall denote the euclidean n-space by R. The elements $x \in R$ will be considered as vectors with $|x|$ the norm of x. The n-cell E and the $(n-1)$-sphere S are defined by the conditions $|x| \leqq 1$, $|x| = 1$.

THEOREM 6. *Let $T: E \to R$ be a continuous multi-valued function such that $T(x) \subset E$ for $x \in S$. If for some coefficient group $G \neq 0$ all the sets $T(x)$ are acyclic then T has a fixed point.*

Proof. Consider the euclidean $2n$-space $R \times R$, compactified to the $2n$-sphere $\overline{R \times R}$ by the addition of a point ∞. In $R \times R$ consider the set Γ_E consisting of those pairs (x, y) satisfying the conditions $x \in E$ and $y \in T(x)$. The subset Γ_S of Γ_E is defined by requiring $x \in S$.

Consider the mappings $r_E : \Gamma_E \to E \times 0$ and $r_S : \Gamma_S \to S \times 0$ defined by $(x, y) \to (x, 0)$. Both r_E and r_S satisfy condition (V) and therefore by Vietoris' theorem

(1) There is an $(n-1)$-dimensional cycle Z in Γ_S such that $Z \sim 0$ in Γ_E but $r_S(Z)$ non ~ 0 on $S \times 0$.

The cycle $r_S(Z)$ links some n-dimensional (integral) cycle on $\overline{0 \times R}$. Since Z and $r_S(Z)$ are homologous (even homotopic) outside of $\overline{0 \times R}$ it follows that

(2) Z links some (integral) cycle on $\overline{0 \times R}$.

Consider the diagonal D consisting of points of the form (x, x). Consider the homotopy,

$$h_t(0, x) = (tx, x), \quad h_t(\infty) = \infty \quad \text{for} \quad 0 \leqq t \leqq 1$$

which deforms $\overline{0 \times R}$ isotopically onto $\bar{D}$.

Suppose first that the path of this homotopy intersects Γ_S. Then $(tx, x) \in \Gamma_S$ or $|tx| = 1$ and $x \in T(tx)$. Since $T(x) \subset E$ for $x \in S$, it follows that $t = 1$ and $x \in T(x)$. This gives a fixed point.

If the path of the homotopy h_t does not intersect Γ_S then by (2) Z links

some cycle of $\bar{D}$. Since $Z \sim 0$ in Γ_E it follows that $\Gamma_E \cap \bar{D} \neq 0$. Hence $\Gamma_E \cap D \neq 0$ and T must have a fixed point.

UNIVERSITY OF MICHIGAN
AND
PRINCETON UNIVERSITY.
SMITH COLLEGE
AND
THE INSTITUTE FOR ADVANCED STUDY.

BIBLIOGRAPHY.

1. Alexandroff, P. and Hopf, H., *Topologie* I, Berlin 1935.
2. Borsuk, K., "Sur le plongement des espaces dans les rétractes absolus," *Fundamenta Mathematicae*, vol. 27 (1936), pp. 239-243.
3. Kakutani, S., "A generalization of Brouwer's fixed point theorem," *Duke Mathematical Journal*, vol. 8 (1941), pp. 457-459.
4. Lefschetz, S., "On the fixed point formula," *Annals of Mathematics*, vol. 38 (1937), pp. 819-822.
5. Lefschetz, S., *Algebraic topology*, New York 1942.
6. Lefschetz, S., "Topics in topology," *Annals of Mathematical Studies* 10, Princeton 1942.
7. Vietoris, L., "Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen," *Mathematische Annalen*, vol. 97 (1927), pp. 454-472.
8. Von Neumann, J. and Morgenstern, O., *Theory of games and economic behavior*, Princeton 1944.
9. Wallace, A. D., "A fixed point theorem for trees," *Bulletin of the American Mathematical Society*, vol. 47 (1941), pp. 757-760.

THE VALUES OF THE NORMS IN ALGEBRAIC NUMBER FIELDS.*

By Aurel Wintner.

1. Let $F = F(n)$, where $n = 1, 2, \cdots$, be a function the values of which are non-negative integers k. Let $\pi_k(m)$ denote the probability that F is in its k-th state when n does not exceed m, that is, let m times $\pi_k(m)$ be the number of those among the first m positive integers n for which $F(n)$ attains the value k. If the limit $\pi_k(\infty)$ exists for every k $(= 0, 1, \cdots)$ and if the sum of the infinite series $\pi_0(\infty) + \pi_1(\infty) + \cdots$ is 1, the function $F(n)$ is said to have an asymptotic distribution function. The latter is represented by the monotone step-function $\alpha(x)$, $-\infty < x < \infty$, which is 0 when $x < 0$ and has the (non-negative) jump $\pi_k(\infty)$ at $x = k$.

The second of the assumptions required for the existence of an asymptotic distribution function is independent of the first. In other words, if the limit $\pi_k(\infty)$ exists for every k, the "total probability" can be distinct from 1 (though not of course greater than 1). In fact, it is easily realized that it is precisely this possibility that is responsible for the type of paradox represented by the so-called Petersburg problem (cf. [7], p. 134 and pp. 220-222).

This possibility is just a manifestation of the fact that, whether a measure or probability be relative or not, Fatou's theorem in Lebesgue's theory is an inequality which cannot in general be replaced by an equality. Correspondingly, even if $F(n)$ has an asymptotic distribution function, and even if there exists an asymptotic mean

$$M(F) = \lim_{m\to\infty} \sum_{n=1}^{m} F(n)/m, \tag{1}$$

all that can be said is that the first moment of the asymptotic distribution function,

$$\int_{-\infty}^{\infty} x\, d\alpha(x), \text{ i.e., } \sum_{k=0}^{\infty} k\,\pi_k(\infty), \tag{2}$$

cannot be less than the value (1). However, if the function $F(n)$ is almost-periodic (B), then not only do both the mean and the asymptotic distribution function exist but, in addition, the mean is equal to the first moment,

* Received March 14, 1945.

$$M(F) = \sum_{k=0}^{\infty} k\pi_k(\infty). \tag{3}$$

Cf. [2], pp. 747-749.

2. With reference to an algebraic number field $\mathfrak{K} = \mathfrak{K}(\theta)$, let $F(n) = F(n; \mathfrak{K})$ denote the number of representations of a positive integer n as norms of integral ideals of $\mathfrak{K}$. In other words, $F(n)$ is the coefficient of $1/n^s$ in the Dirichlet series of Dedekind's $\zeta(s) = \zeta(s; \mathfrak{K})$,

$$\zeta(s) = \Pi\,(1 - (N\mathfrak{p})^{-s})^{-1}, \tag{4}$$

where s is greater than 1 and $\mathfrak{p}$ runs through all prime ideals of $\mathfrak{K}$. If $\mathfrak{K}$ is the rational field, then

$$\zeta(s) = \sum_{n=1}^{\infty} F(n)/n^s \tag{5}$$

becomes Riemann's zeta-function, i. e., $F(n) = 1$ holds for all n. But it turns out that, *except when $\mathfrak{K}$ is the rational field, almost all positive integers n cannot be represented as norms of (integral) ideals of $\mathfrak{K}$, i. e., all but $o(m)$ of the first m terms of the Dedekind zeta-series of $\mathfrak{K}$ are missing.*

In the above terminology, this can be expressed by saying that $F(n)$ possesses an asymptotic distribution function, $\alpha(x)$, since the asymptotic probabilities $\pi_k(\infty)$ of the various states exist and are such as to make

$$\pi_0(\infty) = 1, \quad \text{hence} \quad \pi_1(\infty) = \pi_2(\infty) = \cdots = 0, \tag{6}$$

except when $\mathfrak{K}$ is the rational field. In this exceptional case, (6) must of course be replaced by

$$\pi_1(\infty) = 1, \quad \text{hence} \quad \pi_0(\infty) = 0 \quad \text{and} \quad \pi_2(\infty) = \pi_3(\infty) = \cdots = 0,$$

since $\pi_1(x)$ is 1 for every $x (\geqq 1)$ in this case.

According to Dirichlet-Dedekind (cf., e. g., [3], p. 230), the mean (1) of $F(n)$ exists for every $\mathfrak{K}$. In addition,

$$M(F) \neq 0, \tag{7}$$

since (1) is the residue (at $s = 1$) of the function (4), which has (at $s = 1$) a simple pole. Finally,

$$F(n) = 0, 1, 2, \cdots, \quad \text{hence} \quad F \geqq 0. \tag{8}$$

But the content of the assertion (6) is that "$F(n)$ is 0 almost all of the time". It follows therefore from (8), (7) and (1) that "in the mean, $F(n)$ is very large when it is not 0".

3. During the last decade, various classical results concerning the existence of mean-values of arithmetical functions, and of remainder terms of "explicit sum formulae" in the analytic theory of numbers, have been replaced by results establishing the almost-periodicity (B) of these functions. Such results imply the corresponding classical results, since the existence of a mean-value is necessary for almost-periodicity (B). On the other hand, it is easy to give examples of functions $F(n)$, even of non-negative functions, which are not almost-periodic (B), although the mean-value (1) exists. However, such examples usually depend on artificial constructions and have therefore no arithmetical significance.

Thus, in order to show that the improvements of the classical results in question are not without arithmetical relevance, it is desirable to exhibit a class of functions of *arithmetical* significance which are not almost periodic (B), although their mean-values exist. And it is easy to conclude that the above functions $F(n)$ are of this type, except when $\mathfrak{K}$ is the rational field.

In fact, let us suppose that $F(n)$ is almost-periodic (B). Then (3) is applicable. But the term belonging to $k = 0$ in (3) is 0, since the k-th term of (3) contains the factor k. It follows therefore from (6) that the expression (3) is 0. But this contradicts (7).

Remark. The assertion (6) supplies, for every fixed k, limiting values, $\pi_k(\infty)$, for the probabilities $\pi_k(m)$ as $m \rightarrow \infty$. There arises the finer question as to the asymptotic behavior of the error terms $\pi_k(\infty) - \pi_k(m)$ as $m \rightarrow \infty$, where k is arbitrarily fixed. The answer to this question is known in case $\mathfrak{K}$ is Gauss' quadratic field, $\mathfrak{K}(i)$; cf. [8], pp. 61-66. Actually, the method applied in [8], pp. 61-66, to the Gaussian field could be transferred to the case of every algebraic number field $\mathfrak{K}$ (excepting the trivial case of the rational field), if use is made of those asymptotic results concerning factorizations in $\mathfrak{K}$ which were announced (without proof; cf. pp. 263-265 and p. 537 of Hilbert's *Zahlbericht* [3]) by Kronecker [5] and subsequently proved, in refined formulations, by Frobenius [1]. However, this procedure would involve the same analytical machinery on which the Prime Number Theorem depends. In fact, it would involve extensions of Ikehara's theorem; cf. [8], p. 65 and p. 66. On the other hand, the proof of (6) itself will be "elementary" (in *every* sense customary in the analytic theory of numbers).

4. With reference to a fixed field $\mathfrak{K}$ and to every rational prime p, let $j=j(p)\geqq 1$ denote the number of the distinct prime ideals dividing the principal ideal $[p]$, and let $g_1=g_1(p)\geqq 1,\cdots, g_j=g_j(p)\geqq 1$ be the respective degrees of these $j=j(p)$ prime ideals. Then (4) can be rearranged into

$$(9)\qquad \zeta(s)=\Pi\,(1-p^{-g_1 s})^{-1}\cdots(1-p^{-g_j s})^{-1},$$

where $s>1$. Since $F(n)$ may be defined by identifying (9) with (5), it follows that

$$(10)\qquad F(n_1n_2)=F(n_1)F(n_2)\quad\text{if}\quad (n_1,n_2)=1$$

(i. e., if n_1 and n_2 are relatively prime).

If $e_1=e_1(p)\geqq 1,\cdots, e_j=e_j(p)\geqq 1$ denote the respective multiplicities of the $j=j(p)$ distinct prime ideals occurring in the factorization of the principal ideal $[p]$, then the sum $e_1g_1+\cdots+e_jg_j$ is independent of p, since it is the degree of $\mathfrak{K}$. But it is known that the factorization of $[p]$ into prime ideals contains a multiple factor if and only if the rational prime p which defines $[p]$ is a divisor of the discriminant of $\mathfrak{K}$. In particular, every $e_h=e_h(p)$ is 1 as soon as p^2 exceeds the discriminant of $\mathfrak{K}$. Consequently, the sum of the $j=j(p)$ positive integers $g_h=g_h(p)$ is the degree of $\mathfrak{K}$ as soon as p is large enough.

Since (9) is identical with (5), it follows, in particular, that

$$(11)\qquad F(p^i)\leqq \text{const.}$$

holds for all rational primes p, for all positive integers i and for a certain constant determined by $\mathfrak{K}$ alone.

It also follows that the difference

$$(12)\qquad \sum_{N<x}\frac{1}{N\mathfrak{p}}-\sum_{p<x}\frac{F(p)}{p}$$

tends to a finite limit as $x\to\infty$. In order to see this, it is sufficient to identify both (5) and (9) with (4), and to observe that

$$\sum_p\sum_{m=2}^{\infty}\frac{\text{const.}}{p^m}<\sum_p\frac{\text{const.}}{p^2}<\infty.$$

5. A function $F(n)$ of the positive integer n is called multiplicative if it satisfies condition (10).

If S is a set of distinct positive integers, its characteristic function, $S(n)$, is defined to be the function which is 1 or 0 according as n is or is not in S. A set S having a multiplicative characteristic function is called a multiplicative set.

If $F(n)$ is any multiplicative function, and if S_F denotes the set of those positive integers n at which $F(n)$ is distinct from 0, then S_F is a multiplicative set. In order to see this, it is sufficient to compare condition (10) with the definition of a multiplicative set. In fact, it is clear that a set S is a multiplicative set if, and only if, it has the following property: The product of two integers which are relatively prime is in S or is not in S according as both factors are in S or at least one of them is not in S.

With reference to any multiplicative set S, let S_m denote the number of those of its elements which do not exceed m. Then the ratio S_m/m tends to a limit as $m \to \infty$, and the value of this limit is 0 or positive (but of course not greater than 1) according as the sum of the reciprocal values of those prime numbers which are *not* contained in S is divergent or convergent. This is an immediate consequence of the sieve-process of Eratosthenes; cf., e. g., [8], p. 68.

Since the coefficients of (5) determine a multiplicative function of n, the corresponding set F_S, the n-set on which $F(n)$ does not vanish, is a multiplicative set, and so the ratio S_m/m will tend to 0 or to a positive limit according as the sum of the reciprocal values of those primes p for which $F(p)$ does *not* vanish is divergent or convergent. It follows therefore by subtraction, that (6) is equivalent to the following statement: The sum of the reciprocal values of all primes p satisfying $F(p) = 0$ is divergent,

$$\sum_{F(p)=0} \frac{1}{p} = \infty. \tag{13}$$

Accordingly, everything will be proved if it is verified that (13) is true for every algebraic number field distinct from the rational field. (In the case of the rational field, the sum (13), instead of being divergent, is vacuous, i. e., 0.)

6. According to Mertens' elementary approximation to the Prime Number Theorem, the difference

$$\sum_{p<x} \frac{1}{p} - \log \log x \tag{14}$$

tends to a finite limit as $x \to \infty$. And all the known proofs of this fact (cf., e. g., [4], pp. 22-24) apply, without any change, to the case of any algebraic number field; in the sense that the difference

$$(15) \qquad \sum_{N<x} \frac{1}{N\mathfrak{p}} - \log\log x$$

tends, as $x \to \infty$, to a finite limit in every $\mathfrak{K}$ (for a detailed proof, cf. [6], pp. 150-151).

Since all three differences (12), (14), (15) tend to finite limits as $x \to \infty$, the same is true of the difference

$$\sum_{p<x} \frac{F(p)}{p} - \sum_{p<x} \frac{1}{p}.$$

This means that the infinite series

$$(16) \qquad \sum_{p} \frac{F(p)-1}{p},$$

in which p runs through the (monotone) sequence of *all* rational primes p, is convergent. But every $F(p)$ is a non-negative integer. Hence, the summation in (13) runs over those primes p for which the corresponding term of the series (16) becomes negative. It follows, therefore, from the convergence of (16) that (13) must be true unless the complementary series, i. e., the series

$$(17) \qquad \sum_{F(p)>0} \frac{F(p)-1}{p},$$

is convergent. Consequently, it is sufficient to verify the divergence of the series (17).

Since every $F(p)$ is an integer, every $F(p)-1$ occurring in (17) is either 0 or not less than 1. Hence, in order to assure the divergence of the series (17), it is sufficient to ascertain that

$$(18) \qquad \sum_{F(p)>1} \frac{1}{p} = \infty.$$

But the assertions of Kronecker [5], verified by Frobenius [1], imply the truth of (18) for every algebraic number field $\mathfrak{K}$, except for the rational field.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] G. Frobenius, "Ueber Beziehungen zwischen den Primidealen eines algebraischen Körpers und den Substitutionen seiner Gruppe," *Sitzungsberichte der Akademie der Wissenschaften zu Berlin*, 1896, pp. 689-703.

[2] P. Hartman and A. Wintner, "On the standard deviations of additive arithmetical functions," *American Journal of Mathematics*, vol. 62 (1940), pp. 743-752.

[3] D. Hilbert, "Zahlbericht," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 4 (1894-95 [1897]), pp. 177-546.

[4] A. E. Ingham, *The Distribution of Prime Numbers*, Cambridge Tracts No. 30, 1932.

[5] L. Kronecker, *Werke*, vol. 2, pp. 85-93.

[6] E. Landau, "Ueber die zu einem algebraischen Zahlkörper gehörige Zetafunktion und die Ausdehnung der Tschebyschefschen Primzahltheorie auf das Problem der Verteilung der Primideale," *Journal für die reine und angewandte Mathematik*, vol. 125 (1903), pp. 64-188.

[7] I. Todhunter, *A History of the Mathematical Theory of Probability*, Cambridge and London, 1865.

[8] A. Wintner, *Eratosthenian Averages*, Baltimore, 1943.

THE ASYMPTOTIC NUMBER OF LATIN RECTANGLES.*

By Paul Erdös and Irving Kaplansky

1. Introduction. The problem of enumerating n by k Latin rectangles was solved formally by MacMahon [4] using his operational methods. For $k=3$, more explicit solutions have been given in [1], [2], [3], and [5]. While further exact enumeration seems difficult, it is an easy heuristic conjecture that the number of n by k Latin rectangles is asymptotic to $(n!)^k \exp(-{}_kC_2)$. Because of an error, Jacob [2] was led to deny this conjecture for $k=3$; but Kerawala [3] rectified the error and then verified the conjecture to a high degree of approximation. The first proof for $k=3$ appears to have been given by Riordan [5].

In this paper we shall prove the conjecture not only for k fixed (as $n \to \infty$) but for $k < (\log n)^{3/2-\epsilon}$. As indicated below, a considerably shorter proof could be given for the former case. The additional detail is perhaps justified by (1) the interest attached to an approach to Latin squares ($k=n$), (2) the emergence of further terms of an asymptotic series (4), (3) the fact that $(\log n)^{3/2}$ appears to be a "natural boundary" of the method. (We believe however that the actual break occurs at $k=n^{1/3}$.)

2. Notation. An n by k Latin rectangle L is an array of n rows and k columns, with the integers $1, \cdots, n$ in each row and all distinct integers in each column. Let N be the number of ways of adding a $(k+1)$-st row to L so as to make the augmented array a Latin rectangle. We use the sieve method (method of inclusion and exclusion) to obtain an expression for N. From $n!$, the total number of possible choices for the $(k+1)$-st row, we take away those having a clash with L in a given column—summed over all choices of that column, then reinstate those having clashes in two given columns, etc. The result can be written

$$N = \sum_{r=0}^{n} (-)^r A_r (n-r)! \tag{1}$$

where A_r is the number of ways of choosing r distinct integers in L, no two in the same column. In particular $A_0 = 1$, $A_1 = nk$. To estimate the higher values of A_r we apply the sieve method again. The total number of ways of

* Received November 30, 1945.

selecting r elements of L, not necessarily distinct integers but with no two in the same column, is ${}_nC_r k^r$. This over-estimates A_r; we have to take away those selections which include a specified pair of 1's, 2's, $\cdots$, or n's, then reinstate those which include two pairs, etc. We may write the result

$$(2) \qquad A_r = \sum_s (-)^s B(r, s).$$

Here $B(r, s)$ is precisely defined as follows. Take any s of the $n\, {}_kC_2$ pairs of 1's, $\cdots$, n's which can be formed in L. Suppose that this selection involves in all y elements; y may be as large as $2s$, or as small as the integer for which ${}_yC_2 = s$. Find the number of ways of adjoining $r - y$ further elements, so as to form a set of r elements with no two in the same column. The result of summing over all choices of s pairs is, by definition, $B(r, s)$. We note in particular that

$$(3) \qquad B(r, 0) = {}_nC_r\, k^r$$

$$(4) \qquad B(r, 1) = n\, {}_kC_2\, {}_{n-2}C_{r-2}\, k^{r-2}.$$

The B's may be analyzed further as follows. Let $F(s, t)$ be the number of ways of choosing s pairs of 1's, $\cdots$, n's, which use up t elements in all, and for which no two of the t elements lie in the same column. The number of ways of expanding this selection of t elements to r elements, with no two in the same column, is ${}_{n-t}C_{r-t}\, k^{r-t}$. Hence

$$(5) \qquad B(r, s) = \sum_t F(s, t)\, {}_{n-t}C_{r-t}\, k^{r-t}.$$

It is to be observed that extreme limits for the summation in (5) are given by $t \leqq 2s$ and $s \leqq {}_tC_2$ or, more generously, $\sqrt{s} \leqq t$.

These quantities $F(s, t)$ are the ultimate building blocks from which the exact value of N is constructed. We shall discuss them further in **4.** For the present the following crude inequality will suffice:

$$(6) \qquad \sum_s F(s, t) < n^{t/2} (k^2 t)^{t^2}.$$

The proof of (6) is as follows. The left hand side is just the number of ways of choosing a set of (any number of) pairs which involve in all precisely t elements. In such a choice at most $[t/2]$ distinct integers are permissible, and these may be taken in less than $n^{t/2}$ ways. In all we have at most

${}_tC_2 < t^2$ pairs to dispose of in the selection. For each of these t^2 pairs we have ${}_kC_2\, t/2 < k^2t$ possibilities and hence for all of them at most $(k^2t)^{t^2}$ choices. This establishes (6).

The various quantities defined in this section will be used without further explanation in the remainder of the paper.

3. Proof of the main result. We first prove

THEOREM 1. *If* $k < \log n)^{3/2-\epsilon}$, *then for sufficiently large* n

$$|Ne^k/n! - 1| < n^{-c} \tag{7}$$

where c *is a positive constant depending only on* ϵ.

Proof. Define $A(r,x)$ by

$$A(r,x) = \sum_{s=1}^{x-1} (-)^s B(r,s), \tag{8}$$

where $x = [(\log n)^{1-\epsilon}]$. Then by the sieve's well known property of being alternately in excess and defect we have

$$|A_r - B(r,0) - A(r,x)| \leqq B(r,x). \tag{9}$$

In (1) make the substitution

$$A_r = \{A_r - B(r,0) - A(r,x)\} + B(r,0) + A(r,x)$$

and use (3) and (9). We find

$$|N - \sum_{r=0}^{n} (-)^r\, {}_nC_r\, k^r (n-r)!\,| \leqq |G| + H, \tag{10}$$

where

$$G = \sum_{r=0}^{n} (-)^r A(r,x)(n-r)!, \tag{11}$$

$$H = \sum_{r=0}^{n} B(r,x)(n-r)!. \tag{12}$$

We proceed to study G. With the use of (8) and (5), and an interchange of summation signs, (11) becomes

$$G/n! = \sum_{s=1}^{x-1} (-)^s \sum_t F(s,t) \sum_{r=t}^{n} (-)^r \, {}_{n-t}C_{r-t} \, k^{r-t}/(n)_r$$

where $(n)_r = n(n-1)\cdots(n-r+1)$ is the Jordan factorial notation. The change of variable $r = t + u$ transforms the final sum into

$$(-)^t/(n)_t \sum_{u=0}^{n-t} (-k)^u/u! = (-)^t (e^{-k} - \theta)/(n)_t$$

where θ is the remainder after $n - t$ terms of the series for e^{-k}. Then

$$(13) \qquad |G| \, e^k/n! \leqq \sum_{s=1}^{x-1} \sum_t F(s,t)(1 + \theta e^k)/(n)_t.$$

As noted above, the limits for t lie between $\sqrt{s}$ and $2s$. Hence $t \leqq 2x < 2 \log n$. From this we readily deduce

$$(14) \qquad 1/(n)_t < c_1 n^{-t},$$

$$(15) \qquad \theta e^k < c_2,$$

where c_1, c_2 are absolute constants. From (6), (13), (14), and (15) we obtain

$$|G| \, e^k/n! < c_3 \sum_{t=1}^{2x} (k^2 t)^{t^2}/n^{t/2}$$

with $c_3 = c_1(1 + c_2)$. In the fraction under the summation sign, the logarithms of numerator and denominator are respectively of the orders $t^2 \log\log n$ and $t \log n$. Since $t < 2(\log n)^{1-\epsilon}$, it follows that for large n

$$(k^2 t)^{t^2}/n^{t/2} < n^{-c_4}$$

where c_4 is a positive constant depending only on ϵ. Hence

$$(16) \qquad |G| \, e^k/n! < 2x c_3 n^{-c_4} < n^{-c_5}.$$

We next turn our attention to the term H given by (12). From (5) and an interchange of orders of summation,

$$H/n! = \sum_t F(x,t) \sum_{r=t}^{n} {}_{n-t}C_{r-t} \, k^{r-t}/(n)_r.$$

The final sum is the product of $1/(n)_t$ by a portion of the series for e^k. Hence

4

$$H/n! < e^k \sum_t F(x,t)/(n)_t < c_1 e^k \sum_t (k^2t)^{t^2}/n^{t/2}$$

by (6) and (14). The fraction to be estimated is the same as above but the summation now starts at $\sqrt{x} \geqq c_6(\log n)^{(1-\epsilon)/2}$. It follows that $t \log n \geqq c_6(\log n)^{3/2-\epsilon/2}$, and we are able to swallow up a further term e^{2k} whose logarithm is less than $2(\log n)^{3/2-\epsilon}$. Hence for large n

$$e^{2k}(k^2t)^{t^2}/n^{t/2} < n^{-c_7}$$

and

$$H e^k/n! < 2xc_1 n^{-c_7} < n^{-c_8}. \tag{17}$$

Combining (16), (17), and (10), we obtain (7), for the sum on the left of (10) may run to infinity at a cost of $O(n^{-c})$. This concludes the proof.

(We may note that for the case where k is fixed as $n \to \infty$, the proof could be abridged as follows. We take $x = 1$; then the term G disappears, and an estimate of H is easily obtained from (4).)

From Theorem 1 we readily derive our main result:

THEOREM 2. *Let $f(n,k)$ be the number of n by k Latin rectangles and suppose $k < (\log n)^{3/2-\epsilon}$. Then*

$$f(n,k)(n!)^{-k} \exp({}_kC_2) \to 1 \quad \textit{as} \quad n \to \infty. \tag{18}$$

Proof. From Theorem 1 it follows that $f(n, i+1)$ lies between the limits $f(n,i)n!\,e^{-i}(1 \pm n^{-c})$. Taking the product from $i = 1$ to $k-1$, we find that $f(n,k)$ lies between the limits

$$(n!)^k \exp(-{}_kC_2)(1 \pm n^{-c})^k.$$

Since $(1+n^{-c})^k$ and $(1-n^{-c})^k \to 1$ as $n \to \infty$, we obtain (18).

4. Further terms of the asymptotic series. A more careful argument reveals that the error term in (7) is actually of the order of k^2n^{-1}. By detaching the term $B(r,1)$ as well as $B(r,0)$ in (2), we can reduce the error to the order of k^4n^{-2}. Continuing in this fashion, we may compute successive terms of an asymptotic series. The existence of such a series was conjectured by Jacob [2, 337].

We shall merely sketch the results. Applying (1), (2), and (5) as we did in **3**, we find

$$N/n! = \sum_s (-)^s \sum_t F(s,t)(e^{-k} - \theta)/(n)_t.$$

The term θ may be dropped and we have

$$(19)\qquad Ne^k/n! = 1 - \frac{F(1,2)}{(n)_2} + \frac{F(2,3)}{(n)_3} + \frac{F(2,4)}{(n)_4} - \cdots$$

Thus all that is required is evaluation of the F's. That $F(1,2) = n\,{}_kC_2$ was already implicitly noted in (4). For $F(2,3)$ we observe that not more than one integer may be used, that there are then $n\,{}_kC_3$ choices for the three elements, and 3 choices for the two pairs within them. Hence $F(2,3) = 3n\,{}_kC_3$. Similarly $F(2,4)$ includes the term $3n\,{}_kC_4$, corresponding to the choice of only one integer. If two different integers are taken, there are *ab initio* ${}_nC_2({}_kC_2)^2$ choices; but we must eliminate selections which include two elements in the same column. An application of the sieve process to this last difficulty yields

$$F(2,4) = 3n\,{}_kC_4 + {}_nC_2({}_kC_2)^2 - n\,{}_kC_2(k-1)^2 + X,$$

where X is the number of instances in which integers i, j both occur in two different columns. It is noteworthy that this is the first term which depends upon the particular Latin rectangle to which a $(k+1)$-st row is being added.

A simple argument shows that $X \leqq n\,{}_kC_2(k-1)$, so that $X/(n)_4$ is of order n^{-3} or less, as are all the later terms of (19). Hence we have, correct up to n^{-2}:

$$(20)\qquad \begin{aligned} Ne^k/n! &= 1 - \frac{n\,{}_kC_2}{(n)_2} + \frac{2n\,{}_kC_3}{(n)_3} + \frac{{}_nC_2({}_kC_2)^2}{(n)_4} + \cdots \\ &= 1 - {}_kC_2/n + {}_kC_2(k+4)(3k-7)/12n^2 + \cdots. \end{aligned}$$

By taking the product of the terms (20) from 1 to $k-1$, we obtain the asymptotic series for $f(n,k)$, the number of Latin rectangles:

$$(21)\qquad \begin{aligned} &f(n,k)\,(n!)^{-k}\exp({}_kC_2) \\ &\quad = 1 - {}_kC_3/n + {}_kC_3(k^3 - 3k^2 + 8k - 30)/12n^2 + \cdots. \end{aligned}$$

For $k = 3$, the right side of (21) becomes $1 - 1/n - 1/2n^2 + \cdots$. In the table below we compare this with the exact value given by Kerawala in [3].

n	$1 - 1/n - 1/2n^2$	Exact value of (21)
5	.78	.76995
10	.895	.89560
15	.93111	.93126
20	.94875	.94881
25	.9592	.95923

In attempting to push the asymptotic series still further, we run into the difficulty that terms like X, i. e., terms dependent upon the preceding Latin rectangle, begin to play a rôle in (20). However, it may be that in (21) at least the term in n^{-3} can be obtained without consideration of X, for heuristically it seems likely that the "expectation" of X is $o(n)$.

In conclusion we remark that the form of (21) strongly suggests that at about $k = n^{1/3}$ the expression ceases to be valid. We are unable to prove this rigorously.

STANFORD UNIVERSITY
AND
UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY

1. L. Dulmage, Problem E 650, *American Mathematical Monthly*, vol. 51 (1944), pp. 586-587.

2. S. M. Jacob, "The enumeration of the Latin rectangle of depth three . . .", *Proceedings of the London Mathematical Society*, vol. 31 (1930), pp. 329-354.

3. S. M. Kerawala, "The enumeration of the Latin rectangle of depth three by means of a difference equation," *Bulletin of the Calcutta Mathematical Society*, vol. 33 (1941), pp. 119-127.

4. P. A. MacMahon, *Combinatory Analysis*, vol. I, Cambridge, 1915, pp. 246-263.

5. J. Riordan, "Three-line Latin rectangles," *American Mathematical Monthly*, vol. 51 (1944), pp. 450-452.

A MATRIX DIFFERENTIAL EQUATION OF RICCATI TYPE.*

By William T. Reid

1. **Introduction.** This paper is concerned with the matrix differential equation

$$W' + WA(x) + D(x)W + WB(x)W = C(x), \tag{1.1}$$

where $A(x)$, $B(x)$, $C(x)$ and $D(x)$ are given $n \times n$ square matrices whose elements are continuous functions [1] of the real variable x on the finite and closed interval $ab: a \leqq x \leqq b$. Section **2** of this paper contains generalizations of some well known theorems on the solutions of a single ordinary differential equation of Riccati type. In particular, Theorems 2.2 and 2.3 provide decided extensions of results proved by Whyburn [6] [2] for the matrix differential equation

$$W' + WW = C(x). \tag{1.2}$$

Section **3** is concerned with the relation of the matrix equation (1.1) to a system of $2n$ linear homogeneous differential equations. Finally, in Section **4** it is shown that the analogue of Legendre's differential equation for simple integral problems of the calculus of variations in $(n+1)$-space is a matrix differential equation of the form (1.1).

Throughout this paper capital italic letters denote n-rowed square matrices. In particular, $n \times 1$ rectangular matrices are referred to as vectors. The transpose of a matrix A is indicated by $\tilde{A}$, the reciprocal of a non-singular matrix A by A^{-1}, and, if the elements of A are differentiable functions, the matrix of derivatives is denoted by A'.

2. **Solutions of the matrix differential equation (1.1).** The following preliminary result of Lemma 2.1 has been noted previously by the author [5]

Lemma 2.1. *If the elements of the matrices $H(x)$, $K(x)$ are con-*

* Received August 30, 1945. Presented to the Society, November 25, 1944.

[1] One might allow the elements of these matrices to be merely Lebesgue integrable on ab; in this case, in (1.1) and the subsequently considered matrix differential equations it is understood that a *solution* is a matrix whose elements are absolutely continuous and which satisfies the differential equation almost everywhere on ab.

[2] Numbers in square brackets refer to the bibliography at the end of this paper.

tinuous on ab; then the general solution $T(x)$ of the matrix differential equation

$$T' = H(x)T + TK(x) \tag{2.1}$$

is of the form $T(x) = T_1(x)CT_2(x)$, where $T_1(x)$ is a non-singular (fundamental) solution of $T'_1 = H(x)T_1$, $T_2(x)$ is a non-singular solution of $T'_2 = T_2K(x)$, and C is an arbitrary constant matrix.

COROLLARY 1. *A solution $T(x)$ of (2.1) is of constant rank on ab.*

COROLLARY 2. *If $T_1(x), \cdots, T_{kn+1}(x)$, $1 \leqq k \leqq n$, are solutions of (2.1), then there exist constants $c_1, \cdots, c_{kn+1}$ such that on ab the matrix*

$$T(x) = \sum_{\beta=1}^{kn+1} c_\beta T_\beta(x) \tag{2.2}$$

has constant rank not exceeding $n-k$.

The result of Corollary 2 may be obtained, for example, by choosing the constants c_β so that for $x = a$ the elements of k rows of the matrix (2.2) are equal to zero.

Now if $T_1(x)$ is a non-singular solution of $T'_1 = H(x)T_1$, then $T_2(x) = T_1^{-1}(x)$ is a non-singular solution of $T'_2 = -T_2H(x)$. The following result is then an immediate consequence of Lemma 2.1, and the similarity of $T_1(x)CT_1^{-1}(x)$ to a constant matrix.

COROLLARY 3. *The general solution of the matrix differential equation*

$$T' = H(x)T - TH(x) \tag{2.3}$$

is of the form $T(x) = T_1(x)CT_1^{-1}(x)$, where $T_1(x)$ is a non-singular solution of $T'_1 = H(x)T_1$, and C is an arbitrary constant matrix. For a given solution $T(x)$ of (2.3) the coefficients of the characteristic equation of $T(x)$ are constants.

In the following theorems we shall be concerned with solutions $W(x)$ of the non-linear matrix differential equation (1.1) for which the elements of $W(x)$ are continuous on ab; such solutions will be referred to as continuous solutions.

THEOREM 2.1. *If $W_1(x)$ and $W_2(x)$ are two continuous solutions of (1.1) on ab, then $V(x) = W_2(x) - W_1(x)$ is a continuous solution of the matrix differential equation*

$$(2.4)\qquad V' + V[A(x) + B(x)W_1(x)] + [D(x) + W_1(x)B(x)]V + VB(x)V = 0$$

on ab; moreover, $V(x)$ is of constant rank on this interval.

It may be verified directly that $V = W_2 - W_1$ satisfies (2.4). Now (2.4) may be written as an equation in V of the form (2.1) with

$$H(x) = -[D(x) + W_1(x)B(x) + V(x)B(x)],$$
$$K(x) = -[A(x) + B(x)W_1(x)],$$

and hence $V(x)$ is of constant rank on ab by the above Corollary 1.

THEOREM 2.2. *If $V_1(x)$ and $V_2(x)$ are non-singular continuous solutions of the matrix differential equation*

$$(2.5)\qquad V' + VA(x) + D(x)V + VB(x)V = 0$$

on ab, then $V_2^{-1}(x) - V_1^{-1}(x)$ has constant rank on this interval. Moreover, if $V_1(x), \cdots, V_{kn+2}(x)$, $1 \leqq k \leqq n$, are non-singular continuous solutions of (2.5), then for each value of α, $(\alpha = 1, \cdots, kn+2)$, there exist constants $c_{\beta,\alpha}$, $(\beta = 1, \cdots, kn+2;\ \beta \neq \alpha)$, such that on ab the matrix

$$(2.6)\qquad \sum_{\substack{\beta=1\\ \beta\neq\alpha}}^{kn+2} c_{\beta,\alpha}[V_\alpha^{-1}(x) - V_\beta^{-1}(x)]$$

has constant rank not exceeding $n-k$.

If $V(x)$ is a non-singular continuous solution of (2.5) on ab, then $U(x) = V^{-1}(x)$ is a solution of the non-homogeneous equation

$$(2.7)\qquad U' = A(x)U + UD(x) + B(x)$$

on this interval. Consequently, if $V_1(x)$ and $V_2(x)$ are non-singular continuous solutions of (2.5) on ab, then $T(x) = V_2^{-1}(x) - V_1^{-1}(x)$ is a solution of the homogeneous equation

$$(2.8)\qquad T' = A(x)T + TD(x),$$

and hence, by Corollary 1, the matrix $V_2^{-1}(x) - V_1^{-1}(x)$ is of constant rank on ab. Correspondingly, if $V_1(x), \cdots, V_{kn+2}(x)$, $1 \leqq k \leqq n$, are non-singular continuous solutions of (2.5), then for a fixed value of α the matrices $T_\beta(x) = V_\alpha^{-1}(x) - V_\beta^{-1}(x)$, $(\beta = 1, \cdots, kn+2;\ \beta \neq \alpha)$, are $kn+1$ solutions of (2.8), and by Corollary 2 there exist constants $c_{\beta,\alpha}$ such that on ab the matrix (2.6) has constant rank not exceeding $n-k$.

THEOREM 2.3. *Suppose that*

$$W_\beta(x), \ (\beta = 1, \cdots, kn+3; \ 1 \leqq k \leqq n),$$

are continuous solutions of (1.1) *on ab, and that for a given index* γ *the* $kn+2$ *matrices* $W_\beta(x) - W_\gamma(x)$, $(\beta = 1, \cdots, kn+3; \beta \neq \gamma)$, *are non-singular. Then for each value of* α, $(\alpha = 1, \cdots, kn+3; \alpha \neq \gamma)$, *there exist constants* $c_{\beta,\alpha;\gamma}$ *such that on ab the two matrices*

$$(2.9) \quad \sum_{\substack{\beta=1 \\ \beta\neq\alpha,\gamma}}^{kn+3} c_{\beta,\alpha;\gamma}[W_\beta - W_\alpha][W_\beta - W_\gamma]^{-1}, \quad \sum_{\substack{\beta=1 \\ \beta\neq\alpha,}}^{kn+3} c_{\beta,\alpha;\gamma}[W_\beta - W_\gamma]^{-1}[W_\beta - W_\alpha]$$

have the same constant rank, which does not exceed $n-k$.

In view of Theorem 2.1, each of the matrices

$$V_\beta(x) = W_\beta(x) - W_\gamma(x), \ (\beta = 1, \cdots, kn+3; \ \beta \neq \gamma),$$

is a non-singular continuous solution of

$$(2.10) \qquad V' + V[A(x) + B(x)W_\gamma(x)] \\ + [D(x) + W_\gamma(x)B(x)]V + VB(x)V = 0.$$

The application of Theorem 2.2 to these solutions $V_\beta(x)$ of (2.10) then implies that for a given $\alpha \neq \gamma$ there exist constants $c_{\beta,\alpha} = c_{\beta,\alpha;\gamma}$ such that on ab the matrix

$$(2.11) \qquad \sum_{\substack{\beta=1 \\ \beta\neq\alpha,\gamma}}^{kn+3} c_{\beta,\alpha;\gamma}[V_\alpha^{-1} - V_\beta^{-1}]$$

has constant rank not exceeding $n-k$. The fact that the first matrix of (2.9) has the same rank as (2.11) is an immediate consequence of the relations

$$V_\alpha^{-1} - V_\beta^{-1} = V_\alpha^{-1}[V_\beta - V_\alpha]V_\beta^{-1} = [W_\alpha - W_\gamma]^{-1}[W_\beta - W_\alpha][W_\beta - W_\gamma]^{-1}.$$

Similarly, the relations

$$V_\alpha^{-1} - V_\beta^{-1} = V_\beta^{-1}[V_\beta - V_\alpha]V_\alpha^{-1} = [W_\beta - W_\gamma]^{-1}[W_\beta - W_\alpha][W_\alpha - W_\gamma]^{-1}$$

imply that the second matrix of (2.9) has the same rank as the matrix (2.11).

It is to be remarked that the corresponding result obtained by Whyburn [6] for the equation (1.2) consisted only of the above result for the second matrix (2.9) in the special case $k = n$. It is to be noted, moreover, that if the elements of the matrices $A(x)$, $B(x)$, $C(x)$ and $D(x)$ are real-valued,

and the elements of the solutions $W_\beta(x)$ are also real-valued, then the constants $c_{\beta,\alpha;\gamma}$ of the above theorem may be chosen as real.

In order to derive further results on the form of the general solution of (1. 1), suppose that $W_1(x)$ is a continuous solution of this equation on the interval ab, and let $T(x) = T_\rho(x)$, $(\rho = 3, \cdots, n^2 + 2)$, be n^2 solutions of the homogeneous equation

$$(2.12) \quad T' = [A(x) + B(x)W_1(x)]T + T[D(x) + W_1(x)B(x)]$$

which are linearly independent on ab. In view of Lemma 2. 1 and its Corollary 1, such matrices may be determined by choosing the initial values at x_0 so that the elements of the n^2 matrices $T_\rho(x_0)$ are linearly independent. Now let $U_2(x)$ be a solution of the corresponding non-homogeneous equation

$$(2.13) \quad U'_2 = [A(x) + B(x)W_1(x)]U_2 + U_2[D(x) + W_1(x)B(x)] + B(x)$$

which is such that each of the $n^2 + 1$ matrices

$$U_2(x),\ U_\rho(x) = U_2(x) + T_\rho(x),\ (\rho = 3, \cdots, n^2 + 2),$$

is non-singular on ab. Such a choice is clearly possible, in view of the result of Lemma 2. 1 for the equation (2. 12). Then $W_1(x)$, together with the $n^2 + 1$ matrices

$$(2.14) \qquad W_\beta(x) = W_1(x) + U_\beta^{-1}(x),\ (\beta = 2, \cdots, n^2 + 2),$$

afford $n^2 + 2$ continuous solutions of (1. 1) such that the $n^2 + 1$ matrices

$$V_\beta(x) = W_\beta(x) - W_1(x),\ (\beta = 2, \cdots, n^2 + 2),$$

are non-singular on ab. As

$$V_\beta^{-1}(x) = U_\beta(x),\ (\beta = 2, \cdots, n^2 + 2),$$

from the above construction we have that the matrices

$$V_2^{-1}(x) - V_\rho^{-1}(x) = -T_\rho(x),\ (\rho = 3, \cdots, n^2 + 2),$$

are linearly independent on ab. Now suppose that $W(x)$ is any continuous solution of (1. 1) on ab which is such that the matrix $W(x) - W_1(x)$ is non-singular. If in Theorem 2. 3 we consider $k = n$, $\gamma = 1$, $\alpha = 2$, and identify $W(x)$ with $W_{n^2+3}(x)$, it then follows that the coefficient $c_{n^2+3,2;1}$ in (2. 9) cannot be zero. Consequently, we have the following result as a corollary to Theorem 2. 3.

COROLLARY. *If $W_1(x)$ is a continuous solution of* (1.1) *on the interval ab the matrices $W_\beta(x)$, $(\beta = 2, \cdots, n^2+2)$, determined by* (2.14) *are additional continuous solutions of* (1.1) *such that, if $W(x)$ is an arbitrary continuous solution of this equation on ab for which $W(x) - W_1(x)$ is non-singular, then there exist constants c_ρ, $(\rho = 3, \cdots, n^2+2)$, such that on this interval,*

$$(2.15_1)\qquad [W - W_2][W - W_1]^{-1} = \sum_{\rho=3}^{n^2+2} c_\rho [W_\rho - W_2][W_\rho - W_1]^{-1},$$

$$(2.15_2)\qquad [W - W_1]^{-1}[W - W_2] = \sum_{\rho=3}^{n^2+2} c_\rho [W_\rho - W_1]^{-1}[W_\rho - W_2].$$

3. An associated system of linear differential equations. We shall now consider the relation between the matrix equation (1.1) and the system of $2n$ linear homogeneous differential equations which may be written in vector form as

$$(3.1)\qquad \eta' = A(x)\eta + B(x)\zeta, \qquad \zeta' = C(x)\eta - D(x)\zeta.$$

In (3.1), η and ζ are vectors, that is, $n \times 1$ matrices, with components (η_i) and (ζ_i). The fundamental connection between (3.1) and (1.1) is given by the following theorem.

THEOREM 3.1. *There exists a set of n solutions*

$$\eta_i = \eta_{ij}(x), \qquad \zeta_i = \zeta_{ij}(x), \qquad (j = 1, \cdots, n),$$

of (3.1) *such that the matrix $\| \eta_{ij}(x) \|$ is non-singular on ab if and only if* (1.1) *possesses a continuous solution on ab.*

For suppose that $W(x)$ is a continuous solution of (1.1) on ab, and consider the linear homogeneous matrix differential equation

$$(3.2)\qquad Y' = [A(x) + B(x)W(x)]Y.$$

Let $Y(x) \equiv \| \eta_{ij}(x) \|$ be a non-singular solution of (3.2), and define $Z(x) \equiv \| \zeta_{ij}(x) \|$ as $Z(x) = W(x)Y(x)$. From (3.2) and (1.1) it then follows that

$$(3.3)\qquad Y' = A(x)Y + B(x)Z, \qquad Z' = C(x)Y - D(x)Z,$$

and $\eta_i = \eta_{ij}(x)$, $\zeta_i = \zeta_{ij}(x)$, $(j = 1, \cdots, n)$, define n solutions of (3.1) such that the matrix $\| \eta_{ij}(x) \|$ is non-singular on ab.

On the other hand, if $\eta_i = \eta_{ij}(x)$, $\zeta_i = \zeta_{ij}(x)$, $(j = 1, \cdots, n)$, are solutions of (3.1) such that $\| \eta_{ij}(x) \|$ is non-singular on ab, then

$$Y(x) \equiv \| \eta_{ij}(x) \|, \quad Z(x) = \| \zeta_{ij}(x) \|$$

satisfy (3.2), and it is verified readily that $W(x) = Z(x)Y^{-1}(x)$ is a continuous solution of (1.1) on ab.

4. The Legendre differential equation of the calculus of variations. For a non-parametric fixed end point problem of the calculus of variations the second variation is of the form

$$(4.1) \qquad I_2[\eta] = \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\,dx$$
$$\equiv \int_{x_1}^{x_2} [\tilde{\eta}' R(x)\eta' + 2\tilde{\eta}' Q(x)\eta + \tilde{\eta} P(x)\eta]\,dx,$$

where $R(x)$, $Q(x)$ and $P(x)$ are $n \times n$ square matrices whose elements are real-valued and continuous on x_1x_2, while $R(x)$ and $P(x)$ are symmetric on this interval. In (4.1), η is an admissible variation, that is, an n-dimensional vector whose components $\eta_i(x)$ are continuous and have piece-wise continuous derivatives on x_1x_2, while $\eta_i(x_1) = 0 = \eta_i(x_2)$, $(i = 1, \cdots, n)$.

Along a non-singular extremal E_{12} the matrix $R(x)$ is non-singular, and the canonical form of the accessory (Jacobi) differential equations is

$$(4.2) \qquad \eta' = A(x)\eta + B(x)\zeta, \qquad \zeta' = C(x)\eta - \tilde{A}(x)\zeta,$$

where the coefficients of (4.2) are given by

$$(4.3) \qquad A = -R^{-1}Q, \qquad B = R^{-1}, \qquad C = P - \tilde{Q}R^{-1}Q.$$

That is, the canonical form of the accessory equations is a system (3.1) with $D(x) = \tilde{A}(x)$, and $B(x)$, $C(x)$ symmetric matrices.

Now if E_{12} is a non-singular extremal which has on it no point conjugate to the point 1, there exists a solution $Y(x) \equiv \| \eta_{ij}(x) \|$, $Z(x) \equiv \| \zeta_{ij}(x) \|$ of the system

$$(4.4) \qquad Y' = A(x)Y + B(x)Z, \qquad Z' = C(x)Y - \tilde{A}(x)Z$$

such that $Y(x)$ is non-singular on x_1x_2, while the matrix $\tilde{Y}(x)Z(x)$ is symmetric on this interval (see, for example, Bliss [3], Secs. 12, 36, 39). In view of the non-singularity of $Y(x)$, the symmetry of $\tilde{Y}(x)Z(x)$ is equivalent to the symmetry of the matrix $Z(x)Y^{-1}(x)$. The matrix differential equation of Riccati type associated with (4.4) is

$$(4.5) \qquad W' + WA(x) + \tilde{A}(x)W + WB(x)W - C(x) = 0,$$

which, in view of (4. 3), may be written in terms of the coefficients of (4. 1) as

$$(4.5') \qquad W' + [\tilde{Q}(x) - W]R^{-1}(x)[Q(x) - W] - P(x) = 0.$$

This matrix differential equation may properly be termed the Legendre differential equation for (4. 1), since it is the direct generalization of the equation introduced by Legendre (see Bolza [4], Sec. 9) for the special case $n = 1$. From the symmetry of the matrices $B(x)$ and $C(x)$ it follows, in particular, that if $W(x)$ is a solution of (4. 5) then $\tilde{W}(x)$ is also a solution of this equation. Consequently, if $W(x)$ is a solution of (4. 5) such that at a particular point x_0 the matrix $W(x_0)$ is symmetric, then $W(x)$ is symmetric for all values of x. In view of Theorem 3. 1 we then have that if a non-singular extremal E_{12} has on it no point conjugate to the point 1, then there exists on x_1x_2 a continuous symmetric solution $W(x)$ of the Legendre differential equation (4. 5). For such a solution $W(x)$ we have the following identity

$$(4.6) \qquad 2\omega(x, \eta, \eta') \equiv \tilde{u}R(x)u + (\tilde{\eta}W(x)\eta)',$$

where $u = \eta' + R^{-1}(Q - W)\eta$. One could use the identity (4. 6) to show that if E_{12} is a non-singular extremal along which $I_2[\eta] \geqq 0$ for arbitrary admissible variations η, then $R(x)$ is positive semi-definite, and consequently positive definite on x_1x_2. Such a procedure would not be desirable, however, since without the initial assumption of non-singularity one may show readily that if $I_2[\eta] \geqq 0$ for arbitrary admissible variations then $R(x)$ is positive semi-definite on x_1x_2. In view of the relation between a continuous symmetric solution of (4. 5) and a solution $Y(x)$, $Z(x)$ of (4. 4), the above identity (4. 6) is equivalent to the well known Clebsch (Legendre) transformation of the second variation. In particular, one may state that along a non-singular extremal E_{12} the condition that $I_2[\eta] > 0$ for arbitrary non-identically vanishing admissible variations η is equivalent to the condition that $R(x)$ be positive definite on x_1x_2 and there exists a continuous symmetric solution of (4. 5) on this interval.

The above relations between the canonical form of the accessory differential equations and the Legendre matrix differential equation may be extended readily to problems of Lagrange with fixed end points. Along an extremal E_{12} for such a problem the second variation is of the form (4. 1), while the equations of variation for the side differential equations are of the form

$$(4.7) \qquad \Phi_\beta(x, \eta, \eta') \equiv \phi_{\beta j}(x)\eta'_j + \theta_{\beta j}(x)\eta_j = 0, \qquad (\beta = 1, \cdots, m < n).$$

For simplicity, (4.7) will be written in the matrix form

$$\Phi(x, \eta, \eta') \equiv \phi(x)\eta' + \theta(x)\eta = 0,$$

where $\phi(x) \equiv \| \phi_{\beta j}(x) \|$ and $\theta(x) \equiv \| \theta_{\beta j}(x) \|$ are $m \times n$ matrices. A given extremal E_{12} is termed non-singular if the corresponding $(n+m)$-rowed square matrix

$$\mathfrak{R}(x) \equiv \begin{Vmatrix} R(x) & \tilde{\phi}(x) \\ \phi(x) & 0 \end{Vmatrix} \tag{4.8}$$

is non-singular on x_1x_2. The inverse of such a matrix (4.8) is then of the form

$$\mathfrak{R}^{-1}(x) \equiv \begin{Vmatrix} T(x) & \tilde{\tau}(x) \\ \tau(x) & t(x) \end{Vmatrix}, \tag{4.9}$$

where $T(x)$ and $t(x)$ are symmetric square matrices of orders n and m, respectively, and $\tau(x)$ is an $m \times n$ matrix.

The canonical form of the accessory differential equations is then of the form (4.2), with the coefficient matrices $A(x)$, $B(x)$, $C(x)$ given by

$$A = -(TQ + \tilde{\tau}\theta), \quad B = T, \quad C = P - \tilde{Q}TQ - \tilde{Q}\tilde{\tau}\theta - \tilde{\theta}\tau Q - \tilde{\theta}t\theta.$$

Again, in view of the symmetry of the matrices $B(x)$, $C(x)$, we know that there exists a solution $Y(x)$, $Z(x)$ of (4.4) such that $Y(x)$ is non-singular and $\tilde{Y}(x)Z(x)$ is symmetric on x_1x_2 if and only if there exists a continuous symmetric solution $W(x)$ of the equation (4.5) on this interval. In terms of the coefficients of (4.1) and (4.7) the equation (4.5) now becomes

$$W' + [\tilde{Q}(x) - W, \tilde{\theta}(x)]\mathfrak{R}^{-1}(x)\begin{bmatrix} Q(x) - W \\ \theta(x) \end{bmatrix} - P(x) = 0. \tag{4.5''}$$

Finally, corresponding to (4.6) we have the identity

$$2\omega(x, \eta, \eta') + 2\tilde{\mu}\Phi(x, \eta, \eta') \equiv \tilde{u}R(x)u + (\tilde{\eta}W(x)\eta)', \tag{4.10}$$

where $\mu = [\tau(W - Q) - t\theta]\eta$, and $u = \eta' + [T(Q - W) + \tilde{\tau}\theta]\eta$. The identity (4.10) is equivalent to the Clebsch transformation of the second variation, (see Bliss [1], Sec. 32, and [2], Sec. 22). For a discussion of the relation between the non-existence on E_{12} of a point which is conjugate to 1 and the existence of a solution $Y(x)$, $Z(x)$ of (4.4) of the above described type, the reader is referred to Sec. 23 of Bliss [2], and to the therein referred to papers of Hestenes, Morse and Reid.

NORTHWESTERN UNIVERSITY.

BIBLIOGRAPHY.

[1] G. A. Bliss, "The problem of Lagrange in the calculus of variations," *American Journal of Mathematics*, vol. 52 (1930), pp. 673-744.

[2] G. A. Bliss, *The Problem of Bolza in the Calculus of Variations*, mimeographed notes of lectures delivered at the University of Chicago, Winter Quarter, 1935.

[3] G. A. Bliss, *The Calculus of Variations in Three-Space*, mimeographed notes of revised lectures delivered at the University of Chicago, Winter Quarter, 1938, and in preceding years.

[4] O. Bolza, *Vorlesungen über Variationsrechnung*, Teubner, 1909.

[5] W. T. Reid, "Some remarks on linear differential systems," *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 414-419.

[6] W. M. Whyburn, "Matrix differential equations," *American Journal of Mathematics*, vol. 56 (1934), pp. 587-592.

REVERSIBILITY AND TWO-DIMENSIONAL AIRFOIL THEORY.*

By GARRETT BIRKHOFF.

Introduction. The present paper is concerned with the need for fundamental revisions in the hypotheses underlying mathematical hydrodynamics, with special reference to two-dimensional airfoil theory.

It is first shown, in **2-3**, that *any* reversible theory of lift and drag must be incomplete or grossly incorrect, and remarked that conventional two-dimensional airfoil theory is reversible. This is why it is only applicable to small angles of attack.

The theory of lift, developed in **4**, is shown to be unreliable even for such angles. It leads to absurd conclusions for simple "pathological" shapes (**5**), predicts erroneously the effect on lift of increasing airfoil thickness (**5**), and predicts correctly the effect of camber on lift only at high Reynolds' numbers (**6**). The effect of camber on the stalling angle is predicted by reversibility considerations not mentioned in conventional airfoil theory (**6**).

The question of whether specific reversible hydrodynamical theories are incomplete (or incorrect) is shown to lead to interesting unsolved mathematical problems (**7**). Finally, the reversibility of various other hydrodynamical theories is discussed (**8**).

2. The reversibility paradox. The terminology used below is standard. As regards notation, we shall denote position in space by $x = (x_1, \cdots, x_n)$, time by t, velocity components by u_i, pressure by p, and density by ρ. This notation permits effortless extension to n dimensions of much of the theory of the two-dimensional case.

Definition 1. By the *reverse* of a given flow, described in Lagrangian coordinates, we mean that flow obtained from the given flow by the substitution $t \to -t$. In Eulerian coordinates, this corresponds to the reversal $u_i \to -u_i$ of velocity direction as well as reversal $t \to -t$ of time, but pressure p and density ρ are still preserved at corresponding points $(x_1, \cdots, x_n; t)$ of space-time.

Definition 2. A condition on flows is *reversible* if, whenever it holds for a

* Received October 4, 1945.

flow, it holds for the reverse of that flow. A theory of fluid dynamics is a *reversible theory* if all the conditions which it imposes on flows are reversible.

Actually, all the familiar conditions on flows are reversible, except those which involve viscosity (friction), thermal conduction, diffusion, or shock wave fronts.

Definition 3. A theory of fluid dynamics will be called *incomplete* if its conditions do not determine the steady flow around a body uniquely; *incorrect* if its predictions do not agree closely with experimental fact.

THEOREM 1 (*Reversibility Paradox*[1]). *Any reversible theory of fluid dynamics is either incomplete or grossly incorrect, so far as its predictions of steady state lift and drag are concerned.*

Proof. Such a theory will predict that a steady flow and its reverse will give the same pressure thrust on an obstacle, whereas it is a matter of common experience that a flow and its reverse ordinarily give pressure thrusts in approximately opposite directions.

THEOREM 2 (*Restricted d'Alembert Paradox*). *Any complete reversible theory of fluid dynamics must predict zero lift and drag for steady flow about a body symmetric in a point or in a plane perpendicular to the line of flow.*

Proof. We assume tacitly that the theory is invariant under rigid transformations of space—hence that reflection in a point or in a plane of symmetry replaces each flow permitted by the theory. But with a steady flow, the same effect can be achieved by reversing time, as in Definition 1. Hence, if the theory is complete, the two must be identical, and the pressure distribution must have the symmetry described. The conclusion is now obvious.

Remark 1. In classical hydrodynamics, a stronger result is known:[2] the

---

[1] In special instances, this principle is not new; cf. for example C. Cranz, *Handbuch der Ballistik*, Teubner, 1913, vol. 1, Chap. II, and especially P. Painlevé, *Leçons sur la résistance des fluides*, Paris, 1930, p. 145. Painlevé applies the Reversibility Paradox to classical hydrodynamics, and relates it to the d'Alembert paradox.

Some vaguely related questions are discussed by J. Meixner, "Reversible motions of liquids and gases," *Annalen der Physik*, vol. 41 (1942), pp. 409-425. But in general, physicists seem to ascribe only philosophical interest to the fact that the concept of non-viscous fluid motion is reversible. Thus the fact is not mentioned at all in [1], [2], or [4].

For a direct experimental description of the irreversibility, cf. *N. A. C. A. Tech. Mem.* 1011.

[2] Assuming (1)-(3) of **3**; for the literature, cf. the note by U. Cisotti, *Comptes Rendus*, vol. 178 (1924), p. 1792.

d'Alembert paradox asserts that the predicted lift and drag are zero even for unsymmetrical bodies inclined at any angle to the flow. However, even in this case Theorem 2 is not without interest, because of the extraordinary simplicity of the proof. Existing proofs of the d'Alembert paradox are highly complicated, and open to question on the score of rigor in the compressible case.

Remark 2. The preceding proof breaks down if the restriction to steady flow is omitted. Thus reversible hydrodynamic theory may predict correctly the non-zero drag and lift due to acceleration (virtual mass). Cf. [2], p. 255, p. 419. It also may give a rough idea of the overturn moment to be expected from steady flow.

Remark 3. The d'Alembert paradox shows that the common practice[3] of separating head drag and tail drag in theoretical computations is inadmissible. No matter how long the mid-section of the body, the shape of the tail must theoretically influence the total thrust on the head just as much as the shape of the head itself—since the total thrust is zero.

3. Application to airfoil theory. The classical theory of the steady flow of compressible and incompressible non-viscous fluids (cf. [1], Chaps. IV-VI, [2]) is based on the following mathematical hypotheses. Each mathematical hypothesis, in turn, is based on plausible physical arguments.

There are hypothecated the *equation of continuity*

$$\sum_{k=1}^{n} \partial(\rho u_k)/\partial x_k = 0; \tag{1}$$

the condition that the vector velocity $u(x)$ be tangent to every solid surface; *equation of motion*

$$(1/\rho)\partial p/\partial x_i + \sum_{k=1}^{n} u_k \partial u_i/\partial x_k = 0; \tag{2}$$

the condition of *uniform flow*, which postulates that as $x \to \infty$ the limits of $u_i(x)$, $p(x)$, and $\rho(x)$ exist; the existence of a single-valued *velocity potential*

$$U(x_1, \cdots, x_n) \text{ such that } u_i = \partial U/\partial x_i \text{ for all } i = 1, \cdots, n; \tag{3}$$

the condition that p determine ρ by a thermally controlled *equation of state* $\rho = f(p)$, whose precise form is obtained from physical considerations.

The completeness of these conditions is discussed in **7**. Substituting in

[3] Used in the theory of computing pressure distributions on airship hulls; cf. *N. A. C. A. Report* 516, *Tech. Mem.* 574, also Goldstein, p. 458.

Definitions 1-2 above, it is evident that, insofar as they express the classical theory, *the classical theory of the steady flow of non-viscous fluids is reversible.*

In two-dimensional airfoil theory (cf. [2], Chap. VII, [3], or [4]) the hypothesis (3) of irrotationality in the large is replaced by the slightly weaker assumption

$$\partial u_i/\partial x_k = \partial u_k/\partial x_i \text{ for all } i, k \tag{3'}$$

of irrotationality in the small, and then adding the plausibility hypothesis (4): velocity is finite at the sharp airfoil edge. Again it is evident that *two-dimensional airfoil theory is reversble.*

It is a corollary of this observation and Theorem 1 that two-dimensional airfoil theory can only give correct predictions of lift or drag for a limited range of "angles of attack" (i. e., orientations with respect to the air flow). In fact ([4], pp. 148-151) it fails grossly when the angle of attack exceeds the "stalling angle" of 15°-25°.

Although few numerical calculations have actually been made in the compressible case, we know *a priori* by Theorem 1 that there is no hope that taking account of compressibility will remove the limitation of the applicability of two-dimensional airfoil theory to small angles of attack.

4. Theory of lift: incompressible case. In the case $\rho = \text{const.}$ of an incompressible fluid, there are many numerical calculations based on (1)-(2)-(3′)-(4) which can be compared with experiment.

The predicted drag is still zero (cf. [4], p. 165), and in practice two-dimensional airfoil theory is never used to estimate drag (which depends on "streamlining") for this reason.[4]

However, at small angles of attack, the predicted lift L and lift coefficient C_L (as defined in [2] or [4]) are frequently supposed to represent good estimates. The predicted lift for a given airfoil shape A is most easily found from the Kutta-Joukowsky Theorem, which asserts ([4], p. 163) that the lift on A is the product (for a fluid of density one)

$$L = |u_\infty| \cdot \int \sum_{k=1}^{2} u_k dx_k \tag{5}$$

of the speed $|u_\infty|$ of the airflow by the "circulation" around the airfoil.

[4] Historical remarks. Thus the integral formulas of Newton and Euler (cf. P. Painlevé or C. Cranz, *op. cit.*) are more suggestive, though very misleading. Incidentally, Newton thought that what is usually referred to as the "Newtonian Theory" applied to gases, but not to liquids. Thus he believed the drag of a smooth convex body in a liquid to depend only on its maximum cross-section. (Book II, Lemmas 4, 5; Theorem 29).

But now if A is represented on the complex z-plane, any schlicht conformal transformation

(6) $$w = z + a_1/z + a_2/z^2 + \cdots$$

of the exterior of A will carry potential flows with circulation $U(z)$ into potential flows $U(w(z))$ around the exterior of the image of A, which have the same circulation and the same velocity at infinity. Hence by the Kutta-Joukowsky Theorem, the lift L is the same, and so the lift coefficient C_L (based on the diameter) is inversely proportional to the diameter of the transform. By the fundamental existence and uniqueness theorems of conformal mapping,[5] this determines the C_L of a region of general shape; the argument is due to von Mises [3].

But it is known that any schlicht map

(6′) $$\omega = \zeta + \alpha_2\zeta^2 + \alpha_3\zeta^3 + \cdots$$

of the interior of the unit circle $|\zeta| < 1$ carries it into a region not included in any circle of radius less than one, but including all the circle[6] $|\omega| < .5$.

Under inversion, these inequalities imply that (6) maps the exterior of the unit circle onto the exterior of a region of diameter at *least* two and at *most* four. For a circle, $C_L = 4\pi \sin\theta$ (cf. [4]), where θ is the "angle of attack" between the direction of flow and the radius to the rear stagnation point. Hence for a general profile A

(7) $$C_L = C_L^* \sin(\theta - \theta_0) \quad \text{where} \quad 2\pi \leqq C_L^* \leqq 4\pi;$$

this result is in [3], which is however hardly available to English-speaking readers.

It is another corollary that for equal angles of attack and wind velocities, if A contains a profile A_1, then the predicted lift exerted by A exceeds the predicted lift exerted by[7] A_1. Finally, it follows that the predicted lift coefficient varies continuously with the shape, if the position of the sharp edge is held fixed. (To see this, reduce to the case of the circle; for all shapes between $|z| = 1 - \epsilon$ and $|z| = 1 + \epsilon$, $4\pi(1 - 2\epsilon) \leqq C_L^* \leqq 4\pi$.) Incidentally, C^*_L is proportional to two-dimensional electrostatic capacity, to which these results also apply.

[5] Cf. C. Carathéodory, *Conformal Mapping*, Cambridge University Tracts, p. 70.

[6] The first inequality has an elementary proof; the area of the image circle is by computation $\pi(1 + |\alpha_2|^2 + |\alpha_3|^2 + \cdots) \geqq \pi$. The second is Bieberach's Verzerrungssatz; cf. L. Bieberbach, *Lehrbuch der Funktionentheorie*, Vol. II (1927), Section 9.

[7] This is also a direct corollary of Carleman's Principle of Regional Extension; cf. R. Nevanlinna, *Eindeutige analytische Funktionen*, Springer, 1936.

5. Comparison with experiment: symmetric airfoils. An objective comparison of predicted lift with observed lift brings to light various discrepancies.

In the first place, the importance of the sharp edge is greatly exaggerated. Thus by transforming the circle $|z+99| = 100$ under $z \to z + 1/z$, we get a "Joukowski profile" consisting of a near circle with a small thorn. The predicted lift of this is as great as that of a thin wing having twice the diameter—which is absurd. By transformations $w \to aw + b/w$ $(a, b$ real), we can map the circle with a tiny thorn onto a thin ellipse, with a thorn at the end of a minor axis. Theoretically, tilting the major axis will cause an excess of pressure on the *downstream* side, which is even more absurd.

In the case of conventional airfoil shapes, other discrepancies occur. Here the basic shape is a straight line segment, slightly rounded at the leading edge (cf. [3], [4]). This is varied by introducing added *thickness*, and *camber* (or curvature, usually downward).

For shapes approximating a straight line segment, the conventional theory predicts[8] $C_L = 2\pi \sin \theta$. Observed values at small angles of attack are consistently[9] about 10%-25% less than this.

For symmetric airfoil shapes, the *predicted* C_L^* in $C_L = C_L^* \sin \theta$ *increases* with thickness; the *observed* C_L^* *decreases* with thickness. Thus the sign of the differential change is wrong,[10] and also the absolute discrepancy with theory is even greater than 30%.

Since the sign of the differential change is incorrect, we infer that two-dimensional airfoil theory cannot be expected to predict correctly differential changes in lift due to small changes in profile shape. This does *not* imply, however, that the theory will not give a good idea of the differential changes in flow patterns and pressure distributions due to small changes in profile shape.[11]

---

[8] In the limiting case of a flat plate as noted by Cisotti (*Rendic. Lincei*, 1927), the Kutta-Joukowski Theorem fails. This is because of infinite pressure per unit area at the leading edge, which may be interpreted as giving finite thrust on the leading edge. This Paradox of Cisotti also applies to circular arc profiles.

[9] Cf. [4], pp. 148-151; also *N. A. C. A. Technical Report* 244, Refs. 506, 508; *Report* 628, Figs. 5, 11. Also *Phil. Trans.* 225A (1925), pp. 199-245.

[10] Cf. also Durand, *Aerodynamic Theory*, vol. 2, p. 71. Th. von Kármán suggested to the author the following interpretation. The actual flow differs from the predicted flow in that a thick wake is shed by the downstream rear end of the airfoil. Thus in a certain sense, the effective angle of attack is less than the nominal angle of attack. The force of this argument is greatest for thick airfoils, which have the thickest wakes.

[11] Methods of computing these have been worked out by M. Munk, and by T. Theodorsen; cf. *N. A. C. A. Report* 411. The author would take the statement on p. 10 that "The moments about any required axis may be found" with reservation, as it

6. Cambered airfoils. If θ_0 is the angle between the chord joining the opposite ends of an airfoil and the airflow when $C_L = 0$, the lift coefficient is

$$C_L = C_L{}^* \sin(\theta - \theta_0) \tag{7'}$$

both theoretically and (approximately) experimentally when $\theta - \theta_0$ is small. With a symmetric airfoil, $\theta_0 = 0$ by symmetry.

In the typical case of shapes approximating [8] a circular arc, the theoretical values of $C_L{}^*$, θ_0 are easily found. Under the transformation

$$w = z + 1/z; \quad \text{or} \quad u = x + x/r^2, \quad v = y - y/r^2, \tag{8}$$

the inverse image of a general circle $u^2 + v^2 + 2Cv = 4$ through $(\pm 2, 0)$ is the locus $r^4 + 1 + 2(x^2 - y^2) + 2Cy(r^2 - 1) = 4r^2$.

If $C = -A + 1/A$, this can be factored into

$$(r^2 - 1 - 2Ay)(r^2 - 1 + 2y/A) = 0.$$

Hence (8) maps the two orthogonal circles $r^2 - 2Ay = 1$, $r^2 + 2y/A = 1$ through $(\pm 1, 0)$ into $u^2 + v^2 + 2Cv = 4$. In particular ([4], p. 179), the exterior of $x^2 + y^2 - 2Ay = 1$ is mapped on the exterior of that part of the arc of $u^2 + v^2 + 2Cv = 4$ in the upper half-plane. The "camber" of this arc, which passes through $(0, 2A)$, is defined as the ratio (maximum distance from chord)/(chord length) $= 2A/4 = A/2$.

A flow around the circular arc has finite velocity at the trailing edge $(-2, 0)$ if and only if the flow around the circle $x^2 + y^2 - 2Ay = 1$ has zero velocity at $(-1, 0)$. This is true if the flow is parallel to the diameter through $(-1, 0)$, which passes through $(0, A)$ and hence makes an angle of $\theta_0 = -\operatorname{Tan}^{-1} A$ with the chord. Furthermore, since the circle mapped on the arc has a radius $\sqrt{1 + A^2}$, as compared with 1 for a straight line, we have in (7'), for "camber" $A/2$

$$C_L{}^* = 2\pi\sqrt{1 + A^2}, \qquad \theta_0 = -\operatorname{Tan}^{-1} A. \tag{9}$$

However, the experimental facts are more complex. The original experiments of Eiffel [12] at low Reynolds numbers indicated that down-camber produced a very large increase in $C^*_L = (dC_L/d\theta)\max$; moreover the observed θ_0 was

---

refers to differentials of integrated pressure effects, and these do not seem to be correctly predicted by theory.

[12] Cf. G. Eiffel, *The Resistance of the Air and Aviation*, translated by J. Hunsaker, London and Boston, 1913. Thus (p. 482) with an estimated "camber" $A/2$ of 3/20, an increase in $C_L{}^*$ of about 50% was observed; one of only 4.5% being predicted. Moreover $\theta_0 = -2°$ instead of the predicted $-15°$. Cf. also [4], p. 151, where with an estimated camber of 3/20, also *N. A. C. A. Rep.* 93, Refs. 77, 117.

nearly zero. Later experiments [9] at intermediate Reynolds numbers indicated a smaller but considerable increase in C_L^*, and a larger negative θ_0. The most recent experiments at high Reynolds numbers find, as predicted, a very small increase in C_L^* with down-camber and a change in the angle θ_0 of zero lift quite near to that predicted.[13]

In summary, *the predicted variations in C_L^* and θ_0' due to camber are observed at high Reynolds numbers, but not at low ones.*

The following theoretical explanation of the experimental fact that down-camber increases the effective stalling angle and maximum lift is perhaps also of interest. By reversibility, nature should abhor infinite velocities at the leading edge just as much as at the trailing edge of an airfoil. Physically, high relative velocities at the leading edge will cause separation there, i. e., stalling. But with a circular arc airfoil of camber $\gamma = A/2$, finite velocity at the ends corresponds to having zero velocity at the inverse image points $(\pm 1, 0)$ of the circle mapped on the airfoil by (8). By reversibility, this corresponds to a horizontal wind direction, and hence to a zero angle of attack, or by (9) effective angle of attack $(0 - \theta_0) = \mathrm{Tan}^{-1} A$. Thus *camber γ should not change the stalling angle, should increase the effective stalling angle by* $\mathrm{Tan}^{-1} 2\gamma \approx 2\gamma$, and so (using an empirical value $C_L^* = 4$) *increase the maximum lift by roughly* 8γ.

7. Mathematical completeness. After the discussion of Sections **2-3**, there still remains the academic question of whether classical hydrodynamics and modern airfoil theory are incomplete or incorrect. As this question leads to important unsolved problems in pure mathematics which have not attracted general attention either among pure or applied mathematicians, it seems worth discussing here.

The answer is threefold. In the *incompressible* case, both theories have been proved [14] to be complete, by rigorous mathematical arguments. Actually, for certain shapes in three dimensions, the classical theory is overdetermined; no solution is possible. In the compressible *subsonic* case, it is guessed [15] that

---

[13] Cf. *N. A. C. A. Tech. Reps.* 460, 628 by E. N. Jacobs, where a variable density wind-tunnel at 20 atm. pressure and 70-100 f/s wind speed were used.

[14] For the classical theory, the most general proof is that of L. Lichtenstein, *Hydrodynamik*, p. 422. This applies to non-homogeneous incompressible fluids, and depends on the theory of linear partial differential equations of elliptic type; in the homogeneous case, only potential theory is needed. For modern airfoil theory (homogeneous case), completeness follows from the ref. of footnote 5.

[15] Cf. H. Bateman, "Two-dimensional motion of a compressible fluid," *Proceedings of the Royal Society of London*, 125 A (1929), pp. 598-618, where the difference between the subsonic and supersonic cases is brought out; also Th. von Kármán, "Compressibility effects in aerodynamics," *J. Aer. Sci.*, vol. 8 (1941), pp. 337-356, and the review

the classical theory is determinate or "complete"; at all events, it is equivalent to a non-linear partial differential equation of elliptic type with boundary conditions. In the *supersonic* case, one gets a differential equation of hyperbolic type, and therefore one may guess that there are many solutions. By putting further irreversible conditions on the solution, it may be possible to get a correct complete theory of drag.[16] Little is known about the theory of circulation for compressible fluids.

We thus find several unsolved problems in pure mathematics which have important applications to theoretical hydrodynamics. (1) The uniqueness of the solution of the differential equations with boundary conditions corresponding to subsonic flow, of compressible non-viscous fluids. (2) Existence of a solution to the same. (3) Corresponding problems (with presumably a contrary answer as regards uniqueness) for supersonic flow. (4) A rigorous discussion of the relation betwen the uniformity hypothesis (3) as usually stated, and the analyticity at infinity usually tacitly assumed.[17] It is at this point that the author questions the rigor of the proofs of the d'Alembert paradox for compressible fluids which exist in the literature; the point acquires special importance in view of the fact that the assumption of analyticity at infinity may be too strong to permit solutions.

In the two-dimensional case, one can reduce to linear partial differential equations by use of the hodograph method,[18] but the boundary conditions become very involved. The author proposes to apply this method also to the discussion of the behavior of compressible fluids at infinity and at cusps (which applies to airfoil theory) in a later paper.

8. Other aspects of reversibility. In any discussion of reversibility, it should be recalled that the kinetic theory of gases is reversible in the sense

---

thereof in *Mathematical Reviews* (1942), p. 220. Incidentally, the statement on p. 8, lines 7-10 of von Kármán's "The problem of resistance in compressible fluids," *Reale Accademia d'Italia* (1936-XIV), seems to be contradicted by the example of the Lebesgue spine in potential theory. For the supersonic case, cf. the latter paper, esp. p. 17. The author has interpreted these papers freely, with the idea of expressing what seem to be the most expert guesses by applied mathematicians.

[16] This is essentially what is done in Th. von Kármán and N. B. Moore, "Resistance of slender bodies, etc.," *Transactions of the American Society of Mechanical Engineers* (1932), pp. 303-310.

[17] For rigorous discussions bearing on the incompressible case, cf. O. D. Kellogg, *Potential Theory*, now being reprinted in this country, Chap. X, Section 8, and Chapter VIII, Section 3. The case of two dimensions reduces to complex variable theory.

[18] Cf. Stefan Bergman, *The Hodograph Method in the Theory of Compressible Fluids*, Providence (Brown University), 1942. The author is indebted to Dr. Bergman for many stimulating conversations, and especially for the point made here.

of the dynamics of systems of particles.[19] Thus for a hydrodynamical theory to be irreversible, it must involve *statistical* mixing effects tending to dissipate energy (increase entropy).

Various other theories are reversible as far as general principles of fluid mechanics are concerned, and are irreversible only because of some single special assumption.

This is true of the Helmholtz-Kirchoff theory of wake ([2], Chap. XII), which is reversible except for the empirical postulation of a stagnant wake behind instead of ahead of the obstacle. Also, the von Kármán-Moore approximate theory of drag,[16] for bullets moving at supersonic speeds, is reversible except where they select arbitrarily the rear instead of the forward real portion of the complex solution of a hyperbolic differential equation.

Perhaps some of the confusion which seems to remain[20] in the fundamentals of the theory of resistance to surface waves can be attributed to the difficulty of finding irreversible hypotheses of a general nature.

In concluding, we note the difficulty of constructing any smooth theory which will predict a drag proportional to the square of the velocity. Indeed, the *vector* drag is proportional to u^2 if $u > 0$ and to $-u^2$ if $u < 0$, under such circumstances. Hence we have a *singularity* at $u = 0$, in the sense of analytic functions; this must correspond to a singularity in the theory. In Newton's theory,[4] this is due to a reversal of the face on which wind is supposed to press; in the Helmholtz-Kirchoff theory of wake, to a reversal of the location of wake. In general, the difficulty in continuing a theory through $u = 0$ may be ascribed to the Reversibility Paradox. It is easy to give plausible general reasons why drag should be proportional to u^2, but difficult to explain the reversal in sign of drag with reversal in sign of u, by a single theory.

HARVARD UNIVERSITY.

BIBLIOGRAPHY.

[1] H. Lamb, *Hydrodynamics*, 6th Edition, Cambridge University Press.

[2] L. M. Milne-Thompson, *Theoretical Hydrodynamics*, London, 1938.

[3] R. von Mises, *Zeits, Flugt. u. Motorluftschiffahrt*, 1917, pp. 157-163, and 1920, pp. 67-73 and 87-89.

[4] Prandtl-Tietjens, *Applied Hydro- and Aeromechanics*, McGraw-Hill, 1934.

[19] Cf. for example G. D. Birkhoff, *Dynamical Systems*, New York, 1927, p. 27. In the n-body problem, Lagrangian coordinates (in the hydrodynamical sense) are used.

[20] Cf. [1], Sections 255-256, Section 249. There seem to be various explanations of wave resistance.

A LIMIT THEOREM FOR RANDOM VARIABLES WITH INFINITE MOMENTS.*

By W. Feller.

Let $\{\boldsymbol{X}_k\}$ be an arbitrary sequence of mutually independent random variables and $\{a_n\}$ a monotonic numerical sequence. As usual, we put

$$S_n = X_1 + \cdots + X_n. \tag{1}$$

We consider the

Event $\mathfrak{L}$: "*The inequality*

$$| S_n | > a_n \tag{2}$$

takes place for infinitely many n."

According to the familiar "one or naught law," the probability of $\mathfrak{L}$ can be only zero or one. For the case where the $\boldsymbol{X}_k$ are individually bounded, an extensive theory has been developed and a recent refinement of Kolmogoroff's law of the iterated logarithm [1] enables us to decide in any special case whether the probability is zero or one. Now this theory depends essentially on the central limit theorem. As soon as we leave the domain of applicability of the central limit theorem we find ourselves on practically unknown terrain; the problems receive an entirely new aspect and no systematic tools have as yet been developed for treating the theory. Outside of the theory of the iterated logarithm only one result seems to be known. The following theorem treats a very special case and is given in a form whose probability meaning is not readily intelligible. Its interest is nevertheless considerable inasmuch as it shows the radical change in the character of the limit theorems caused by the absence of finite moments.

Theorem (*P. Lévy–J. Marcinkiewicz*).[2] *Suppose that one has uniformly for large x and all k*

---

* Received November 26, 1945.

[1] W. Feller, "The general form of the so-called law of the iterated logarithm," *Transactions of the American Mathematical Society*, vol. 54 (1943), pp. 373-402.

[2] P. Lévy, "Sur les séries dont les termes sont des variables éventuelles indépendantes," *Studia Mathematica*, vol. 3 (1931), pp. 117-155; J. Marcinkiewicz, "Quelques théorèmes de la théorie des probabilités," *Travaux de la Société des Sciences et des*

$$(3) \qquad cx^{-\alpha} < \Pr\{|\mathbf{X}_k| > x\} < Cx^{-\alpha}$$

where α, c and C are positive constants. Let $\lambda(t)$ be an increasing function such that $\lambda(2t)/\lambda(t) \to 1$ as $t \to \infty$. For the special sequence

$$(4) \qquad a_n = \{n \log n\, \lambda(\log n)\}^{1/\alpha}$$

and $0 < \alpha < 1$ the event $\mathcal{L}$ has probability one (zero) if the series

$$(5) \qquad \sum\{n\lambda(n)\}^{-1}$$

diverges (converges); the theorem remains true also for $1 \leq \alpha < 2$ provided that $E(\mathbf{X}_k) = 0$ (in case $\alpha = 1$ the Cauchy principal value of the expectation is meant).

This theorem was proved by P. Lévy in the case $0 < \alpha < 1$ using the theory of stable distributions. The method does not work for $\alpha \geq 1$. Marcinkiewicz's proof is of an elementary nature. The striking difference between the P. Lévy–Marcinkiewicz theorem and the law of the iterated logarithm becomes apparent if one notices that according to the former any sequence $\{a_n\}$ exhibits the same character as $\{Ma_n\}$, where $M > 0$ is an arbitrary constant. With the iterated logarithm all that can be said is that $\{\phi_n\}$ and $\{\phi_n \pm M/\phi_n\}$ belong to the same class. In particular, $\{\eta \log\log n\}^{1/2}$ would belong to the upper class if $\eta > 2$, to the lower if $\eta < 2$.

Consider now the

Event $\mathcal{L}^*$: "The inequality

$$(6) \qquad |\mathbf{X}_n| > a_n$$

takes place for infinitely many n."

A simple computation shows that, in the cases where the P. Lévy-Marcinkiewicz theorem applies, the events $\mathcal{L}$ and $\mathcal{L}^*$ have the same probability. In other words: if the conditions (3) and (4) are satisfied, *the asymptotic behavior of the sums $\{\mathbf{S}_n\}$ is entirely determined by the last terms $\mathbf{X}_n$: as far as maxima are concerned, $\mathbf{S}_{n-1}$ can be neglected in comparison with $\mathbf{X}_n$.* Considered against the subtle background of the iterated logarithm, this behavior is very crude indeed. We shall see that it is typical for the case of infinite variance and also that surprisingly simple analytic methods suffice to treat this case. (The theory becomes the simpler the fewer moments are finite.)

Lettres de Wilno, Classe des Sciences Mathématiques et Naturelles, vol. 13 (1939), pp. 1-13.

For simplicity we shall consider only the case where all $\mathbf{X}_k$ have the same distribution function

$$\Pr\{\mathbf{X}_k \leq x\} = V(x). \tag{7}$$

This condition can trivially be relaxed in the direction of (3) where it is only required that the distribution functions of the $\mathbf{X}_k$ should not differ too much from a given distribution function. However, the methods of this paper do not apply to the general case of distribution functions varying with k. The fact that absolute values are introduced in (2) is formally another restriction: however, it simplifies the formulation and it is not difficult to separate the cases $\mathbf{S}_n > a_n$ and $\mathbf{S}_n < -a_n$.

THEOREM 1. *Suppose that for some* $0 < \delta < 1$

$$\int_{-\infty}^{+\infty} |x|^{1+\delta}\, dV(x) = \infty \tag{8}$$

but that the first moment exists and

$$\int_{-\infty}^{+\infty} x\,dV(x) = 0. \tag{9}$$

For the particular sequences $a_n = n^{1/1+\delta}$ *and* $a_n = n$ *the event* $\mathfrak{L}$ *has probability one and zero, respectively. For any sequence* $\{a_n\}$ *for which there exists an* ϵ *with* $0 \leq \epsilon < 1$ *such that*[3]

$$a_n n^{-1/(1+\epsilon)} \uparrow \qquad a_n/n \downarrow \tag{10}$$

the probability of $\mathfrak{L}$ *is one or zero according as the series*

$$\sum \int_{|x| \geq a_n} dV(x) \tag{11}$$

diverges or converges.

---

[3] The restriction (10) seems so natural and mild that no effort has been made to remove it. Actually, the proof requires much less than (10), namely

$$n/a_n < \text{Const.}\ (n + \nu/a_{n+\nu}), \qquad \nu > 0 \tag{10a}$$

and

$$\sum_{\nu=n}^{\infty} a_\nu^{-2} = O(n a_n^{-2}). \tag{10b}$$

The last condition is certainly satisfied if, for example, $\liminf (a_{2n}^2/a_n^2) > 2$.

THEOREM 2. *If*

$$\int_{-\infty}^{+\infty} |x| \, dV(x) = \infty, \tag{12}$$

and $a_n = n$, the event $\mathfrak{L}$ has probability one. For any sequence $\{a_n\}$ with

$$a_n/n \uparrow \tag{13}$$

the probability of $\mathfrak{L}$ is one or zero according as (11) *diverges or converges.*

The statement of either theorem can be reformulated to the effect that *the probability of $\mathfrak{L}$ is the same as that of $\mathfrak{L}^*$.*

Proof. If (11) diverges, the implication is trivial. It suffices therefore to assume that (11) converges. Put

$$\mu_k = \int_{|x|<a_k} x dV(x) \tag{14}$$

and

$$X'_k = \begin{matrix} X_k - \mu_k & \text{if} & |X_k| < a_k \\ 0 & \text{if} & |X_k| \geq a_k \end{matrix}. \tag{15}$$

The probability that $X_k \neq X'_k + \mu_k$ is given by the general term of (11). As this series converges, we have with probability one

$$S_n - \sum_{k=1}^{n} (X'_k + \mu_k) = O(1) = o(a_n). \tag{16}$$

In order to prove that, with probability one,

$$\sum_{k=1}^{n} X'_k = o(a_n) \tag{17}$$

it suffices according to a frequently used trick[4] to show that with certainty

$$\sum (1/a_k) X'_k < \infty. \tag{18}$$

Now X'_k has vanishing expectation, and according to a theorem of Khintchine and Kolmogoroff it suffices to prove that

---

[4] The method of proving (17) from (18), applying Kronecker's theorem, is due to H. Rademacher ("Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen," *Mathematische Annalen*, vol. 87 (1922), pp. 112-138). Its usefulness in proving general probability limit theorems has been demonstrated by A. Zygmund (in an article, written in Polish, in *Mathesis Polska*, vol. 8 (1933), pp. 76-87, where the real variable treatment of the theory of probability was outlined). Subsequently, the method has been extensively used in the Polish probability literature and is by now a familiar tool.

(19) $$\sum 1/a_k^2 \int_{|x|<a_k} (x-\mu_k)^2 dV(x) < \infty.$$

Now, putting $a_0 = 0$,

(20) $$\sum_k 1/a_k^2 \int_{|x|<a_k} x^2 dV(x) = \sum_k 1/a_k^2 \sum_{i=1}^{k} \int_{a_{i-1}\leq|x|<a_i} x^2 dV(x)$$
$$= \sum_i \int_{a_{i-1}\leq|x|<a_i} x^2 dV(x) \sum_{k=i}^{\infty} 1/a_k^2.$$

The first of the conditions (10) implies that

$$\sum_{k=i}^{\infty} 1/a_k^2 < i^{2/(1+\epsilon)}/a_i^2 \sum_{k=i}^{\infty} 1/k^{2/(1+\epsilon)} < 3/(1-\epsilon)\cdot i/a_i^2.$$

Therefore the series (20) is less than

$$3/(1-\epsilon)\sum_i i \int_{a_{i-1}\leq|x|<a_i} dV(x) = 3/(1-\epsilon)\sum_i \int_{|x|\geq a_i} dV(x),$$

and the latter series converges by assumption. As the terms of (19) are not exceeded by those of (20), the convergence of (19), and therefore the certainty of (18), have been established.

In view of (16) and (17) it remains only to prove that

(21) $$\sum_{k=1}^{n} \mu_k = o(a_n).$$

We consider first the case of Theorem 1. It follows then from (9) that, for an arbitrary integer N and $n \geq N$,

(22) $$|1/a_n \sum_{k=1}^{n} \mu_k| \leqq O(N/a_n) + 1/a_n \sum_{k=N}^{n} \int_{|x|\geq a_k} |x|\, dV(x)$$
$$\leqq O(N/a_n) + n/a_n \int_{|x|\geq a_n} |x|\, dV(x) + 1/a_n \sum_{k=N}^{n} \int_{a_k\leq|x|<a_n} |x|\, dV(x).$$

Now, using the second condition (10),

(23) $$n/a_n \int_{|x|\geq a_n} |x|\, dV(x) \leq \sum_{i=n}^{\infty} i/a_i \int_{a_i\leq|x|<a_{i+1}} |x|\, dV(x) \leq 2\sum_{i=n}^{\infty} i \int_{a_i\leq|x|<a_{i+1}} dV(x)$$
$$= 2n \int_{|x|\geq a_n} dV(x) + 2\sum_{j=n}^{\infty} \int_{|x|\geq a_j} dV(x).$$

The last series converges by assumption; the first term to the right tends to zero since the integral is the general term of a convergent series with decreasing terms. As for the last expression in (22)

$$1/a_n \sum_{k=N}^{n} \int_{a_k \leq |x| < a_n} |x| \, dV(x) \leq \sum_{k=N}^{n} \int_{|x| \geq a_k} dV(x),$$

and the last series becomes arbitrarily small for N sufficiently large.

It remains to prove (21) for the case of Theorem 2. Then the series (11) cannot converge if a_n/n remains bounded. It follows from (13) that

$$\begin{aligned} \left| 1/a_n \sum_{k=1}^{n} \mu_k \right| &\leq 1/a_n \sum_{k=1}^{n} \int_{|x| < a_k} |x| \, dV(x) \\ &= O(Na_N/a_n) + n/a_n \int_{a_N \leq |x| < a_n} |x| \, dV(x) \\ &\leq O(Na_N/a_n) + \sum_{s=N}^{n} s/a_s \int_{a_{s-1} \leq |x| < a_s} |x| \, dX(x). \end{aligned}$$

That this tends to zero has already been shown in the second part of (23).

CORNELL UNIVERSITY,
ITHACA, N. Y.

SCALE HYPERSURFACES FOR CONFORMAL-EUCLIDEAN SPACE.*

By YUNG-CHOW WONG [1]

1. Introduction. We shall give in this paper generalizations to n-space of some of the results obtained recently by Kasner and DeCicco (3) [2] for the scale curves in conformal maps of a surface upon a plane. It will be observed that this subject is closely connected with the subject of the isoparametric hypersurfaces [3] of Levi-Civita (4) and Segre (6) and incidentally connected with that of the subprojective Riemannian space of Kagan (2) and Schapiro (5). For notation and convention we shall follow Eisenhart (1), and in particular, we shall confine ourselves to real variables and real functions.

The fundamental form

$$ds^2 = e^{2\sigma}(dx_1{}^2 + \cdots + dx_n{}^2), \qquad \sigma = \sigma(x_1, \cdots, x_n), \tag{1.1}$$

represents a conformal-Euclidean n-space C_n, conformable to the Euclidean n-space R_n.[4] The scalar curvature of C_n with fundamental form (1. 1) is (Eisenhart 1, p. 90)

$$R = (n-1)e^{-2\sigma}[2\Delta_2\sigma + (n-2)\Delta_1\sigma], \tag{1.2}$$

where

$$\Delta_1\sigma = \sum_i (\sigma_{,i})^2, \qquad \Delta_2\sigma = \sum_i \sigma_{,ii}, \qquad (i = 1, \cdots, n), \tag{1.3}$$

are the first and second differential parameters with respect to R_n; the indices after the comma indicate partial differentiation.

The hypersurfaces $\sigma = \text{const.}$ in R_n are the *scale hypersurfaces* in the mapping of C_n on R_n.

Any simple family of hypersurfaces $f = \text{const.}$ in R_n is called *quasi-isothermal* if it represents the scale hypersurfaces of a conformal map of some

* Received May 13, 1945.

[1] Harrison Research Fellow at the University of Pennsylvania.

[2] References quoted are listed at the end of this paper.

[3] We remark that the subject of isoparametric hypersurfaces in a space of constant curvature has been investigated by E. Cartan in recent years.

[4] It is found convenient here to write the rectangular Cartesian coordinates in R_n sometimes as $x^1, \cdots, x^n$ and sometimes as $x_1, \cdots, x_n$.

C_n on R_n such that the scalar curvature of C_n is constant over each of the scale hypersurfaces.

The family of hypersurfaces $f = \text{const.}$ can represent the scale hypersurfaces of that class of C_n for which $\sigma = \sigma(f)$. The scalar curvature for this class of C_n is

$$(1.4) \qquad R = (n-1)e^{-2\sigma}\{2\sigma'\Delta_2 f + [2\sigma'' + (n-2)(\sigma')^2]\Delta_1 f\}.$$

Thus, $f = \text{const.}$ is quasi-isothermal if and only if

$$(1.5) \qquad 2\sigma'\Delta_2 f + [2\sigma'' + (n-2)(\sigma')^2]\Delta_1 f = \text{function of } f.$$

Let $F = F(f)$, then $\Delta_2 F = F'\Delta_2 f + F''\Delta_1 f$. Therefore, if we choose F so that

$$(1.6) \qquad \sigma''/\sigma' + \tfrac{1}{2}(n-2)\sigma' = F''/F',$$

then condition (1.5) becomes

$$(1.7) \qquad \Delta_2 F = \text{function of } F.$$

Hence, $f = \text{const.}$ is a family of quasi-isothermal hypersurfaces if and only if a function F of f exists such that (1.7) is satisfied. In particular, if F is any function satisfying this condition, then $F = \text{const.}$ is a family of quasi-isothermal hypersurfaces. The scale function σ of the corresponding mapping is given by (1.6).

We know that if F satisfies

$$(1.8) \qquad \Delta_1 F = \text{function of } F,$$

the hypersurfaces $F = \text{const.}$ are parallel hypersurfaces. Hypersurfaces $F = \text{const.}$ for which both (1.7) and (1.8) are satisfied have been called *isoparametric hypersurfaces* by Levi-Civita (4) and Segre (6), who proved that

All (real) isoparametric hypersurfaces in R_n ($n \geq 2$) are one of the following three types: (a) parallel hyperplanes, (b) concentric hyperspheres, (c)[5] *generalized coaxial cylinders of rotation (i. e., hypersurfaces which are generated from a pencil of concentric hyperspheres in an r-plane ($r \leq n-1$) by means of the ∞^{n-r-1} translations contained in R_n and orthogonal to the r-plane).*

For $n = 2$ this is essentially Kasner and DeCicco's Theorems 2 and 3.

---

[5] We note that (b) may be considered as an extreme case of (c) when $r = n$.

We shall now proceed to prove the following generalizations of Kasner and DeCicco's Theorems 4 and 5:

THEOREM 1. 1. *A family of ∞^1 hyperplanes is a family of quasi-isothermal hypersurfaces if and only if it is a pencil of hyperplanes. The fundamental form of the corresponding C_n is reducible to*

$$ds^2 = J(x_1)(dx_1^2 + \cdots + dx_n^2), \tag{1.9}$$

where J may be constant or not; in the latter case the pencil of hyperplanes must be parallel hyperplanes.

THEOREM 1. 2. *A family of ∞^1 generalized cylinders of rotation which are generated from a family of ∞^1 hyperspheres in an r-plane is a family of quasi-isothermal hypersurfaces if and only if it is (a) a pencil of coaxial cylinders for $n > 2$, or (b) a pencil of circles for $n = 2$. The fundamental form of the corresponding C_n is reducible to*

$$(1.10)_a \qquad ds^2 = J(x_1^2 + \cdots + x_r^2)(dx_1^2 + \cdots + dx_n^2) \qquad (0 < r \leq n),$$

or

$$(1.10)_b \qquad ds^2 = J(x_1^2 + x_2^2)(dx_1^2 + dx_2^2),$$

respectively, where in $(1.10)_b$ $J \neq$ const. or $=$ const. according as the pencil of circles is concentric or not.

COROLLARY 1. 2. *For $n > 2$, a family of ∞^1 hyperspheres is a family of quasi-isothermal hypersurfaces if and only if it is a pencil of concentric hyperspheres. The fundamental form of the corresponding C_n is reducible to*

$$ds^2 = J(x_1^2 + \cdots + x_n^2)(dx_1^2 + \cdots + dx_n^2). \tag{1.11}$$

By definition a *subprojective* Riemannian n-space is a Riemannian space which can be so mapped on R_n that the images of the geodesics are curves lying in planes which pass through a fixed point. Schapiro (5) proved that the C_n's with fundamental forms (1.9) and (1.11) are two of the three and only classes of subprojective Riemannian space. Theorems 1.1 and 1.2 therefore furnish a new characterization of these two classes of subprojective Riemannian space. Schapiro also proved that the C_n with fundamental form $(1.10)_a$ can be so mapped on R_n that the images of its geodesics are curves lying on $(r + 1)$-planes.

2. Proof of Theorem 1. 1. Let $F =$ const. be a family of (non-isotropic) hyperplanes in R_n with rectangular Cartesian coördinates x^h

$(h, i, j, k = 1, \cdots, n)$. Then there exist some definite functions $a_h(F)$, $a_0(F)$ of F such that the equation

$$a_h(F)x^h + a_0(F) = 0 \tag{2.1}$$

is satisfied identically. Differentiations give

$$F_{,i} = -\lambda a_i \equiv -\frac{a_i}{a_h' x^h + a_0'},$$

$$\Delta_2 F = \Sigma F_{,ii} = 2\lambda^2 \Sigma a_i a_i' - \lambda^3 (a_h'' x^h + a_0'') \Sigma a_i^2,$$

$$(\Delta_2 F)_{,j} = [-4\lambda^3 \Sigma a_i a_i' + 3\lambda^4 (a_h'' x^h + a_0'') \Sigma a_i^2] a_j' - \lambda^3 (\Sigma a_i^2) a_j'' + A F_{,j},$$

where A is some function of x^h. From this it follows that the condition (1.7), namely,

$$(\Delta_2 F)_{,[j} F_{,k]} = (\Delta_2 F)_{,j} F_{,k} - (\Delta_2 F)_{,k} F_{,j} = 0,$$

is

$$\{-(\Sigma a_i^2) a_{[j}'' + [-4\Sigma a_i a_i' + 3\lambda (a_h'' x^h + a_0'') \Sigma a_i^2] a_{[j}'\} a_{k]} = 0,$$

i.e., since $\Sigma a_i^2 \neq 0$,

$$a_j'' + B a_j' + C a_j = 0, \tag{2.2}$$

where C is some function of x^h and

$$B = 4 \frac{\Sigma a_i a_i'}{\Sigma a_i^2} - 3 \frac{a_h'' x^h + a_0''}{a_h' x^h + a_0'}. \tag{2.3}$$

Let us consider (2.2) as n linear equations in B and C. Two cases arise according as only one or more than one of the equations in (2.2) are independent. For the former case, we have

$$a_j = b_j p, \qquad a_j' = b_j p', \qquad a_j'' = b_j p'', \tag{2.4}$$

where $p = p(F)$, $b_j = b_j(x^h)$. From (2.4) it follows that $b_j = b_j(F) = \text{const.}$ The family of hyperplanes (2.1) now becomes

$$p(F)(b_h x^h) + a_0(F) = 0,$$

which is a *pencil* of *parallel* hyperplanes. Therefore we may suppose that $p = 1$, so that $a_j = b_j = \text{const.}$ and equations (2.2) are satisfied by $C = 0$. Then we have from (2.1) that

$$\sigma = \sigma(f), \qquad f = b_h x^h, \qquad \Delta_1 f = \Sigma b_i^2, \qquad \Delta_2 f = 0.$$

Therefore equation (1.5) is satisfied. The corresponding C_n has the funda-

mental form $J(b_h x^h)(dx_1{}^2 + \cdots + dx_n{}^2)$, which can of course be reduced to $J(x_1)(dx_1{}^2 + \cdots + dx_n{}^2)$ by a suitable change of coördinates.

We now suppose that in the equations (2.2) in B and C, there are at least two independent equations. Then since $a_j = a_j(F)$, both B and C are functions of F. We have therefore from (2.3)

$$\frac{a_h{}'' x^h + a_0{}''}{a_h{}' x_h + a_0{}'} = s(F), \text{ say.} \tag{2.5}$$

Moreover, since the a_j, as functions of the independent variable F, are now solutions of the linear homogeneous differential equation (2.2), there exist two functions $p(F)$ and $q(F)$ and some constants b_j and c_j such that

$$a_j = b_j p(F) + c_j q(F). \tag{2.6}$$

It is readily seen that because there are more than one independent equations in (2.2), the p and q and also the b_j and c_j are not proportional.

Using (2.6) in (2.1) and (2.5), we have

$$\begin{aligned} &(b_h x^h)p + (c_h x^h)q + a_0 = 0, \\ &(b_h x^h)(p'' - sp') + (c_h x^h)(q'' - sq') + a_0{}'' - sa_0{}' = 0. \end{aligned} \tag{2.7}$$

These two equations must be dependent; otherwise, both $b_h x^h$ and $c_h x^h$ would be functions of F and therefore they could differ only by a constant factor, i. e., the b_j and c_j would be proportional.

This being the case, we have

$$p'' - sp' = rp, \qquad q'' - sq' = rq, \qquad a_0{}'' - sa_0{}' = ra_0, \tag{2.8}$$

where $r = r(F)$. From these it follows that

$$a_0 = b_0 p(F) + c_0 q(F), \qquad (b_0, c_0 \text{ constant}). \tag{2.9}$$

In consequence of (2.6) and (2.9), equation (2.1) can be written as

$$(b_h x^h + b_0)p(F) + (c_h x^h + c_0)q(F) = 0, \tag{2.10}$$

which represents a *pencil* of *non-parallel* hyperplanes.

We now prove that the C_n corresponding to this quasi-isothermal family, which consists of a pencil of non-parallel hyperplanes, is itself Euclidean. From (2.10) we have

$$\sigma = \sigma(f), \qquad f = \frac{b_h x^h + b_0}{c_h x^h + c_0},$$

$$\Delta_1 f = \nu^2(\Sigma b_h{}^2 - 2f\Sigma b_h c_h + f^2 \Sigma c_h{}^2), \qquad \Delta_2 f = 2\nu^2(-\Sigma b_h c_h + f\Sigma c_h{}^2),$$

where $\nu = 1/(c_h x^h + c_0)$. Substitution of these in (1.4) and (1.5) gives

$$R = \nu^2 \cdot (\text{a function of } f) = \text{function of } f.$$

From this it follows that $R = 0$; for if $R \neq 0$, the b_j would be proportional to c_j. Therefore the C_n is an R_n.

3. Proof of Theorem 1.2. Consider a family of ∞^1 generalized cylinders of rotation which are generated from a family of ∞^1 hyperspheres in a fixed r-plane in R_n $(r \leq n)$. We may suppose, after a suitable change of rectangular Cartesian coördinates, that this r-plane is $x_{r+1} = 0, \cdots, x_n = 0$. Then, if $F = \text{const.}$ is this family of cylinders of rotation, there exist some definite functions $a_h(F)$, $a_0(F)$ of F such that the equation

$$(3.1) \qquad \Sigma x_h^2 = 2[a_h(F)x^h + a_0(F)] \qquad (h, i, j, k = 1, \cdots, r)$$

is satisfied identically. Equation (3.1) can be written

$$(3.2) \qquad \Sigma(x_h - a_h)^2 = r^2 \equiv \Sigma a_h^2 + 2a_0.$$

Differentiations of (3.1) give

$$F_{,i} = \lambda(x_i - a_i) \equiv \frac{x_i - a_i}{a_h' x^h + a_0'},$$

$$\Delta_2 F = \Sigma F_{,ii} = n\lambda - 2\lambda^2 \Sigma a_h'(x_h - a_h) - \lambda^3 r^2 (a_h'' x^h + a_0''),$$

$$(\Delta_2 F)_{,j} = -\lambda^3 r^2 a_j'' + [-(n+2)\lambda^2 + 4\lambda^3 \Sigma a_h'(x_h - a_h) + 3\lambda^4 r^2 (a_h'' x^h + a_0'')]a_j' + A F_{,j},$$

where A is some function of x^h. The condition for $\Delta_2 F$ to be a function of F: $(\Delta F)_{,[j} F_{,k]} = 0$ is therefore

$$(3.3) \qquad a_j'' + B a_j' = C(x_j - a_j),$$

where C is some function of x^h and

$$(3.4) \qquad B = (n+2)\lambda^{-1} r^{-2} - 4r^{-2} \Sigma a_h'(x_h - a_h) - 3\lambda(a_h'' x^h + a_0'').$$

We consider first the case $C = 0$, and then show that the case $C \neq 0$ leads to a contradiction. When $C = 0$, (3.3) is of the form $a_j'' + B a_j' = 0$. It follows from this that $B = B(F)$ and then that

$$(3.5)_1 \qquad a_j = b_j p(F) + c_j,$$

where the b's (not all zero) and c's are constants. Furthermore, we now have from (3.4) after simplification that

$$(n-2) a_h' x^h - 3 \frac{a_h'' x^h + a_0''}{a_h' x^h + a_0'} r^2 = \text{function of } F.$$

When we substitute the values a_j from $(3.5)_1$, this becomes

$$(3.6) \qquad (n-2)(b_h x^h) p' - 3 \frac{(b_h x^h) p'' + a_0''}{(b_h x^h) p' + a_0'} r^2 = \text{function of } F.$$

If $(n-2) p' \neq 0$, this equation would demand that $b_h x^h$ be a function of F, which is contradictory to the hypothesis that $F = \text{const.}$ is a family of cylinders of rotation. Therefore we must have $(n-2) p' = 0$.

If $p' = 0$, (3.6) is satisfied and $(3.5)_1$ may be supposed to reduce to

$$(3.7) \qquad a_j = c_j = \text{const.}$$

This, together with (3.1), shows that we have a *pencil* of *coaxial* cylinders.

If $p' \neq 0$, then $n = 2$. And since $b_h x^h \neq$ function of F, equation (3.6) requires that

$$a_0'' p' - a_0' p'' = 0,$$

so that

$$(3.5)_2 \qquad a_0 = b_0 p(F) + c_0 \qquad (b_0, c_0 \text{ constant}).$$

On account of this and $(3.5)_1$, equation (3.1) represents a pencil of cylinders.

Hence we have two cases: (3.7) or (3.5) with $p' \neq 0$. The former case can happen for any $n (> 1)$, while the latter case can happen only for $n = 2$, which is the case of Kasner and DeCicco.

For the case (3.7), equation (3.1) becomes

$$\Sigma x_h^2 = 2 c_h x^h + a_0(F).$$

Consequently,

$$\sigma = \sigma(f), \quad f = \Sigma x_h^2 - 2 c_h x^h, \quad \Delta_1 f = 4(f + \Sigma c_h^2), \quad \Delta_2 f = 2n.$$

Condition (1.5) is therefore satisfied, and the fundamental form of the corresponding C_n is reducible to

$$ds^2 = J(x_1^2 + \cdots + x_r^2)(dx_1^2 + \cdots + dx_n^2).$$

For the case (3.5), equation (3.1) becomes

$$\Sigma x_h^2 = 2(b_h x^h + b_0)p(F) + 2(c_h x^h + c_0).$$

Therefore $\sigma = \sigma(f)$ and

$$f = \frac{\Sigma x_h^2 - 2(c_h x^h + c_0)}{b_h x^h + b_0} \equiv \frac{\Sigma y_h^2 + c}{b_h y^h + b},$$

where b and c are some constants. From this we have, writing $\nu^{-1} = b_h x^h + b_0 \equiv b_h y^h + b$,

$$\Delta_1 f = \nu^2[4(bf - c) + f^2 \Sigma b_h^2],$$

$$\Delta_2 f = 2n\nu - 4\nu^2 b_h y^h + 2\nu^2(\Sigma b_h^2)(\Sigma y_h^2 + c),$$

the latter of which becomes, for $n = 2$,

$$\Delta_2 f = \nu^2[4b + 2f(\Sigma b_h^2)].$$

Therefore from (1.4) and (1.5) we have

$$R = \nu^2 \cdot (\text{a function of } f) = \text{function of } f.$$

R must be zero; otherwise, the family of quasi-isothermal hypersurfaces $f =$ const. would be $\nu^{-1} = b_h x^h + b_0 =$ const., contradictory to hypothesis. Therefore C_n is a Euclidean space.

We now suppose that the C in (3.3) be not zero. If the a_j'' are all zero, (3.3) becomes $Ba_j' = C(x_j - a_j)$. Squaring this and then summing over j, we have $B^2 \Sigma a_h'^2 = C^2 r^2$, showing that the preceding equation can be written $x_j =$ function of F. From this it follows that the family of quasi-isothermal hypersurfaces $F =$ const. is $x_j =$ const., contradictory to our hypothesis. Therefore a_j'' are not all zero. Let us now multiply (3.3) by a_j'', a_j', $x_j - a_j$ respectively and then sum each of the results over j, we have

$$\begin{aligned} \Sigma a_h''^2 + B\Sigma a_h' a_h'' &= C\Sigma a_h''(x_h - a_h), \\ \Sigma a_h' a_h'' + B\Sigma a_h'^2 &= C\Sigma a_h'(x_h - a_h), \\ \Sigma a_h''(x_h - a_h) + B\Sigma a_h'(x_h - a_h) &= Cr^2. \end{aligned} \tag{3.8}$$

These are three equations in C, $X = \Sigma a_h'(x_h - a_h)$, $Y = \Sigma a_h''(x_h - a_h)$, with functions of F as coefficients. We shall now prove that these equations can be

solved for C, X, Y as definite functions of F, and consequently, we would have a contradiction with (3.1).

Eliminating B and C from (3.8) and C from the last two equations of (3.8), we have, respectively,

$$(3.9)\qquad (\Sigma a_h'^2)Y^2 - 2(\Sigma a_h'a_h'')XY + (\Sigma a_h''^2)X^2 + r^2[(\Sigma a_h'a_h'')^2 - (\Sigma a_h'^2)(\Sigma a_h''^2)] = 0,$$

and

$$r^2\Sigma a_h'a_h'' + Br^2\Sigma a_h'^2 = X(Y + BX).$$

The last equation is, on substitution for the value of B from (3.4),

$$(3.10)\qquad r^2\Sigma a_h'a_h'' - XY + (\Sigma a_h'^2 - r^{-2}X^2) \times \left[(n-2)X + (n+2)(\Sigma a_h a_h' + a_0') - 3r^2\frac{Y + \Sigma a_h a_h'' + a_0''}{X + \Sigma a_h a_h' + a_0'}\right] = 0.$$

Writing this for brevity as

$$(3.10')\qquad \alpha - XY + (\beta - r^{-2}X^2)\left[(n-2)X + (n+2)\gamma - 3r^2\frac{Y+\delta}{X-\gamma}\right] = 0,$$

and solving for Y, we have

$$Y = \frac{(n-2)r^{-2}X^4 + 2n\gamma r^{-2}X^3 + \cdots}{2X^2 - \gamma X - 3\beta r^2},$$

where the unwritten terms are of the second or lower degree in X. When this value of Y is substituted in (3.9), the latter becomes a polynomial equation in X:

$$(3.11)\qquad \begin{aligned} &(\Sigma a_h'^2)[(n-2)r^{-2}X^4 + 2n\gamma r^{-2}X^3 + \cdots]^2 \\ &-2(\Sigma a_h'a_h'')[(n-2)r^{-2}X^4 + 2n\gamma r^{-2}X^3 + \cdots](2X^2 - \gamma X - 3\beta r^2)X \\ &+ (\Sigma a_h''^2)(2X^2 - \gamma X - 3\beta r^2)^2X^2 \\ &+ r^2[(\Sigma a_h'a_h'')^2 - (\Sigma a_h'^2)(\Sigma a_h''^2)](2X^2 - \gamma X - 3\beta r^2)^2 = 0. \end{aligned}$$

If $n \neq 2$, the coefficient of X^8 in (3.11) is $(n-2)^2r^{-4}\Sigma a_h'^2$, which is not zero since the a_h' are not all zero.

If $n = 2$, the coefficient of X^6 in (3.11) is

$$(4\gamma r^{-2})^2\Sigma a_h'^2 - 4(\Sigma a_h'a_h'')4\gamma r^{-2} + 4\Sigma a_h''^2 = 4\Sigma(a_h'' - 2\gamma r^{-2}a_h')^2 = 4\Sigma(a_h'' - 2r'r^{-1}a_h')^2;$$

the last equality follows immediately from the definition of r and γ. This coefficient is zero if and only if

$$a_j'' - 2r'r^{-1}a_j' = 0.$$

Comparing this with (3.3) we have equations of the form $Ea_j' = C(x_j - a_j)$. This, as we saw before, would lead to a contradiction.

Therefore in both cases equation (3.11) requires that X be a function of F. Hence the case $C \neq 0$ is impossible, and the proof of Theorem 1.2 has been completed.

THE UNIVERSITY OF PENNSYLVANIA.

REFERENCES.

1. Eisenhart, L. P., *Riemannian Geometry* (1926), Princeton.
2. Kagan, B., "Über eine Erweiterung des Begriffes vom projecktiven Räume und dem zugehörigen Absolut," *Abhandlung des Seminars fur Vektor- und Tensor-analysis* I (1933, Moskau), pp. 12-101.
3. Kasner E. and DeCicco, J., "Geometry of scale curves in conformal maps," *American Journal of Mathematics*, vol. 67 (1945), pp. 157-166.
4. Levi-Civita, T., "Famiglie di superficie isoparametriche nell' ordinario spazio euclideo," *Atti Accad. naz. Lincei, Rend.*, (6), vol. 26 (1937), pp. 355-362.
5. Schapiro, H., "Über die Metrik der subprojecktiven Räume," *Abhandlung des Seminars für Vecktor- und Tensoranalysis* I (1933, Moskau), pp. 102-124.
6. Segre, B., "Famiglie di ipersuperficie isoparametriche neglie spazi euclidei ad un qualunque numero di demensioni," *Atti Accad. naz. Lincei, Rend.*, (6), vol. 27 (1938), pp. 203-207.

A FACTORIZATION OF THE DENSITIES OF THE IDEALS IN ALGEBRAIC NUMBER FIELDS.*

By Aurel Wintner.

Introduction. Let $\mathfrak{K}$ be an algebraic number field, $\zeta(s;\mathfrak{K})$ its zeta-function and $C = C(\mathfrak{K}) > 0$ the residue of this zeta-function (at $s = 1$). According to the fundamental limit-theorem of Dirichlet-Dedekind (cf., e. g., [1], pp. 142-149), the (integral) ideals of $\mathfrak{K}$ have a finite and non-vanishing density, that is, the number of the ideals having a norm not exceeding x is asymptotically proportional to x, as $x \rightarrow \infty$. If Dirichlet's elementary Abelian lemma (cf., e. g., [1], pp. 152-154) is applied to the Dirichlet series of $\zeta(s;\mathfrak{K})$, it becomes evident that the numerical value of the asymptotic density must be the residue C, provided that the existence of this density is granted. What is not evident is the existence of this density. In fact, when Dirichlet and Dedekind developed their theory of unities, their main, or rather sole, purpose was a proof for the existence of an asymptotic factor of proportionality (for historical references cf. [7]). However, as was shown in [7], this existence theorem can today be proved in a manner which completely avoids the theory or the existence of unities.

The positive results of the present paper (some of them will be negative) supply for the asymptotic density a peculiar evaluation, rather than an existence proof. While the classical evaluation involves such data as the regulator and the numbers of real and complex *unities in the field*, the evaluation which will result only contains *data depending directly on the laws of factorization of the integral ideals.* In particular, the residue C will appear as a product extended over the sequence of all rational primes.

This evaluation of the asymptotic density is not difficult to prove; it is "elementary" in the technical sense of the analytic theory of ideals. That it does not seem to have been observed before, may be explained by the fact that it is fully disguised in the case of the rational field, since a_p then becomes 0 in every factor of the infinite product $\Pi(1 + a_p)$ representing the residue (which is 1 in case of the rational field). Incidentally, this trivial case presents the only field for which the product evaluation is *absolutely* convergent. The result is as follows:

* Received December 10, 1945.

With reference to an arbitrary algebraic number field, let

$$(1) \qquad j = j(p) \geqq 1,$$

where $p = 2, 3, \cdots$, *denote the number of the distinct prime ideals which divide the principal ideal* $[p]$, *and let*

$$(2) \qquad g_1 = g_1(p) \geqq 1, \cdots, g_j = g_{j(p)}(p) \geqq 1$$

be the respective degrees of these j *prime ideals. Then the product*

$$(3) \qquad \prod_p (1 - p^{-1})(1 - p^{-g_1})^{-1}(1 - p^{-g_2})^{-1} \cdots (1 - p^{-g_j})^{-1},$$

where p *runs through the sequence of all rational primes in increasing order, is convergent and its value represents the asymptotic density of the integral ideals in the field.*

The proof of the representation (3) of the asymptotic density will depend on (i) the extension to algebraic number fields of Mertens' elementary approximation to the prime number theorem, and (ii) a particular criterion assuring the legitimacy of a formal Eulerian factorization.

As to (i), it is clear, and well-known, that any of the proofs of Mertens' theorem concerning the rational field (in particular, Mertens' own proof and Hardy's "Tauberian" proof; cf. [3], pp. 22, where further references are given) can directly be transcribed to the case of an arbitrary algebraic number field. However, since all these proofs depend, very explicitly, on Stirling's theorem and break down, therefore, in more general situations, it seemed to be worth proving a more general fact, which is the content of the first of the assertions of the theorem italicized below. What concerns (ii), that is, the legitimacy of the formal Eulerian derangement, assertion (ii) of the same theorem will become applicable. Finally, the negative result, (iii), of the theorem will show that the legitimacy of the derangement in question is by no means automatic.

An appendix considers the nature of *analytical limitations imposed on the zeta-function* $\zeta(s; \mathfrak{K})$ *by the laws of factorization in the field* $\mathfrak{K}$, that is, by the specifically *arithmetical* character of the data (2), (3) of $\mathfrak{K}$.

1. By a multiplicative function $F(n)$ of the positive integer n is meant a sequence $F(1), F(2), \cdots$ satisfying $F(nm) = F(n)F(m)$ whenever n and m are relatively prime. Thus, if the trivial case $F(1) = F(2) = \cdots = 0$

is excluded, the function $F(n)$ is uniquely determined by an arbitrary assignment of its values attained when n is a prime power, $n = p^k$, where $p = 2, 3, \cdots$ and $k > 0$ (if $k = 0$, then $F(1) = 1$, since $F(1) \neq 0$).

The general theorem, referred to above, which does not assume the positiveness of the coefficients (and, in case of positive coefficients, is not restricted to those, very explicit, situations in which Stirling's theorem is applicable) runs as follows:

Let $F(n)$ be a multiplicative function for which the Dirichlet series

$$f(s) = \sum_{n=1}^{\infty} F(n)/n^s \tag{4}$$

is absolutely convergent in the half-plane $\sigma > 1$ and represents there a function acquiring a simple pole at the point $s = 1$ (except for a vicinity of this point, the function $f(s)$ need not remain regular along the line $\sigma = 1$). Then

(i) *if the absolute value of $F(p^k)$ is less than a constant multiple of $p^{\theta(k-1)}$ for some fixed $\theta < 1$ and for every prime power p^k, the series*

$$\sum_p \frac{F(p) - 1}{p} \tag{5}$$

is convergent; in addition,

(ii) *under the assumption of (i), the infinite product*

$$\prod_p (1 - p^{-1}) (1 + \sum_{k=1}^{\infty} F(p^k) p^{-k}) \tag{6}$$

is convergent and its value is the residue of $f(s)$ at $s = 1$; however,

(iii) *if the common assumption of (i) and (ii) is satisfied for every $k \neq 2$ but is relaxed from $F(p^2) = O(p^\theta)$ to $F(p) = O(p)$ for $k = 2$, then the assertions of (ii) become false, not only because (6) may become divergent, but also because (6) can converge to a value distinct from the residue; even though the assumptions made before (i) are satisfied.*

It is understood that p in (5) and (6) is supposed to run through the sequence of all primes in *increasing* order. This proviso is necessary, since the convergence of (5) and (6) is not in general absolute.

2. The following proof of (i) will depend on M. Riesz's extension (to Dirichlet series) of Fatou's theorem (on power series). Since this extension can further be generalized so as to involve just a Fourier condition, rather

than regular-analyticity, near the point $t = 0$ on the boundary line $s = 1 + it$ (cf. [4]), the full force of the assumption, according to which the function (4), where $\sigma > 1$ in $s = \sigma + it$, is the sum of const./$(s - 1)$ and of a function which is *regular-analytic at* $s = 1$, will not be needed. The particular case sufficient for the proof of the above formulation states that, if a Dirichlet series

$$\sum_{n=1}^{\infty} a(n)/n^s \tag{7}$$

converges in the half-plane $\sigma > 1$ to a function which remains regular-analytic at the point $s = 1$, then the (trivial necessary) condition

$$\sum_{n=1}^{x} a(n) = o(x) \qquad (x \to \infty) \tag{8}$$

is sufficient for the convergence of the series (7) at the point $s = 1$.

First, it is clear from the assumption $\theta < 1$ of (i) that (4) possesses the absolutely convergent Eulerian factorization

$$f(s) = \prod_p \left(1 + \sum_{k=1}^{\infty} F(p^k)/p^{ks}\right) \tag{9}$$

in the half-plane $\sigma > 1$ (in fact, even $\theta \leqq 1$ is sufficient to this end). If this is applied to the case $F(1) = F(2) = \cdots = 1$, in which $f(s)$ becomes Riemann's $\zeta(s)$, it follows, by subtraction from the logarithm of (9) itself, that

$$\log \frac{f(s)}{\zeta(s)} = \sum_p \frac{F(p) - 1}{p^s} - \frac{1}{2}\{\cdots\} + \cdots \tag{10}$$

holds if σ is sufficiently large (the logarithm refers to the determination which tends to 0 as $\sigma \to \infty$). If σ is sufficiently large, then, by absolute convergence, the expression on the right of (10) can be rearranged into a Dirichlet series (7) in which

$$a(n) = 0 \quad \text{unless} \quad n = p^k. \tag{11}$$

It is clear from the assumption of (i) that each of the remaining coefficients, $a(p^k)$, of the Dirichlet series of the logarithm (10) has an absolute value less than a constant multiple of the $(\theta + \epsilon)$-th power of p^{k-1}, for every fixed $\epsilon > 0$. Since $\theta < 1$, it can be assumed, by choosing a θ somewhat greater than the given θ, that $a(p^k)$ is majorized by a constant multiple of $p^{\theta(k-1)}$, where $\theta < 1$. If this is compared with (11), it follows that the Dirichlet series (7) is absolutely convergent in the half-plane $\sigma > 1$, and that the sum

of those of its terms which do not belong to primes (i.e., in which $n = p^k$, where $k > 1$) is absolutely convergent at the point $s = 1$.

On the other hand, since (7) is identified with (10), the sum of those terms of (7) which do belong to primes is identical with the first series on the right of (10), a series which becomes the series (5) at the point $s = 1$. But the logarithm on the left of (10) remains regular at $s = 1$, since the Dirichlet series (4), where $\sigma > 1$, is supposed to represent a function having a simple pole at $s = 1$. Consequently, the convergence of the series (5) will be proved if it is shown that (8) is satisfied in the present case.

Since $c(p^k)$ is majorized by a constant multiple of $p^{\theta(k-1)}$, it is seen from (11) that (8) will be ascertained if it is verified that the sum

$$\sum_{p^k < x} p^{\theta(k-1)}$$

is $o(x)$ as $x \to \infty$. In order to verify this estimate, let the latter sum be rearranged into

$$\sum_{k=1}^{\infty} \sum_{p < x^{1/k}} p^{\theta(k-1)},$$

where p is the summation index of the interior sum. Since there are just $O(\log x)$ values of k for which $x^{1/k}$ exceeds at least one p, the upper limit, ∞, of the exterior summation can be replaced by $O(\log x)$. On the other hand, since the number of primes not exceeding N is less than a constant multiple of $N/\log N$ (Chebyshev), the number of terms in the k-th interior sum does not exceed a constant multiple of $kx^{1/k}/\log x$. Finally, the greatest term of the k-th interior sum is less than the $\theta(k-1)$-th power of $x^{1/k}$. Consequently, the sum (8) is majorized by a constant multiple of

$$\sum_{k=1}^{O(\log x)} k x^{\theta(k-1)/k} x^{1/k}/\log x.$$

Clearly, the first factor, k, of the k-th term of this majorant can be omitted if ϑ is replaced by a somewhat greater θ. Hence, all that remains to be shown is that, as $x \to \infty$, the estimate

$$\sum_{k=1}^{O(\log x)} x^{\theta(k-1)/k} x^{1/k}/\log x = o(x)$$

holds for every fixed $\theta < 1$. But this is obvious. In fact, if θ were 1, the last sum would be

$$\sum_{k=1}^{O(\log x)} x^{(k-1)/k} x^{1/k}/\log x = x \sum_{k=1}^{O(\log x)} 1/\log x,$$

which is $xO(\log x)/\log x = O(x)$. And this O becomes an o if $\theta = 1$ is replaced by $\theta < 1$.

This completes the proof of (i).

4. The identity (10) resulted from a simultaneous logarithmization of (9) and of the corresponding factorization of the Riemann zeta-function, namely, of the infinite product

$$f(s)/\zeta(s) = \prod_p (1 - p^{-s})(1 + \sum_{k=1}^{\infty} F(p^k) p^{-ks}),$$

where $\sigma > 1$. Let the factors of this infinite product be denoted by $1 + c_p(s)$. Then

$$f(s)/\zeta(s) = \prod_p (1 + c_p(s)), \tag{12}$$

where $\sigma > 1$ and

$$c_p(s) = \alpha_p(s) + \beta_p(s), \tag{13}$$

if $\alpha_p(s)$ and $\beta_p(s)$ are abbreviations for

$$\alpha_p(s) = (F(p) - 1)/p^s \tag{14}$$

and

$$\beta_p(s) = \sum_{k=2}^{\infty} (F(p^k) - F(p^{k-1}))/p^{ks}. \tag{15}$$

Since $(s-1)\zeta(s) \to 1$ as $s \to 1$, the assertion of (ii) is equivalent to the statement that (12) remains valid at the point $s = 1$, if the value attained at $s = 1$ by the quotient on the left of (12) is meant to be the residue of $f(s)$ at $s = 1$. In other words, (ii) will be proved if it is shown that the infinite product (12) is convergent on the *closed* half-line $s \geqq 1$ and represents there a function which is *continuous* (at $s = 1$). In view of (iii), the second of these assertions is independent of the first.

Since the case $k = 1$ of the assumption of (i) means that $F(p)$ is a bounded function of p, it is clear from (14) that

$$\sum_p \operatorname*{Max}_{1 \leqq s} |\alpha_p(s)|^2 < \infty. \tag{16}$$

Similarly, the series (15) is majorized by a constant multiple of

$$\sum_{k=2}^{\infty} 2p^{\theta(k-1)}/p^{ks} \leqq 2 \sum_{k=2}^{\infty} p^{\theta(k-1)-k},$$

if $s \geqq 1$. As $p \to \infty$, the sum of the last series remains less than a constant multiple of its first term, which is $p^{\theta-2}$. Since $\theta - 2 < -1$, it follows that $\Sigma \mid \beta_p(s) \mid$ is majorized for $s \geqq 1$ by a convergent numerical series, and so

$$\sum_p \operatorname*{Max}_{1 \leqq s} \mid \beta_p(s) \mid < \infty. \tag{17}$$

In particular

$$\sum_p \operatorname*{Max}_{1 \leqq s} \mid \beta_p(s) \mid^2 < \infty. \tag{18}$$

According to a standard convergence criterion of Cauchy, an infinite product $\Pi(1 + c_p)$ satisfying $\Sigma \mid c_p \mid^2 < \infty$ is convergent if and only if the series Σc_p is, and a corresponding criterion holds for uniform convergence. But (18), (16) and (13) imply, by Schwarz's inequality, that the series $\Sigma \mid c_p(s) \mid^2$ is uniformly convergent for $s \geqq 1$. Hence, the product (12) is uniformly convergent for $s \geqq 1$ if and only if the same is true of the series $\Sigma c_p(s)$. In view of (13) and (17), this will be the case if and only if the series $\Sigma \alpha_p(s)$ is uniformly convergent for $s \geqq 1$. Finally, (14) shows that $\Sigma \alpha_p(s)$ is a Dirichlet series which at $s = 1$ becomes the series (5), the convergence of which is assured by (i). Since a Dirichlet series which is convergent at $s = 1$ must be uniformly convergent for $s \geqq 1$ (Abel-Jensen), it follows that $\Sigma c_p(s)$ is uniformly convergent for $s \geqq 1$.

Consequently, the product (12) is uniformly convergent for $s \geqq 1$. This is slightly more than what was needed for the completion of the proof of (ii), which was seen to depend on whether or not the product (12) converges for $s \geqq 1$ to a continuous function.

5. What concerns (iii), the possibility of a divergent product (6) is obvious under the assumptions of (iii). The remaining statement of (iii) is that the product (12) may converge on the closed half-line $s \geqq 1$ to a function having a discontinuity of the first kind at $s = 1$, if the assumptions of (iii) are satisfied.

An example proving this possibility results by choosing the multiplicative function $F(n)$ as follows: $F(p^k) = p^{\frac{1}{2}k}$ or $F(p^k) = 0$ according as k is even or odd. Then all assumptions of (iii) are satisfied. Furthermore, since

$$1 + \sum_{k=1}^{\infty} F(p^k) p^{-k} = 1 + \sum_{k=1}^{\infty} p^k p^{-2k} = \sum_{k=0}^{\infty} p^{-k},$$

every factor of the product (6) is 1. Hence, all that remains to be shown is that the function $f(s)$ has at $s = 1$ a simple pole the residue of which is distinct from 1. But (9) shows that, if $\sigma > 1$,

$$f(s)=\prod_p (1+\sum_{k=1}^{\infty} p^k/p^{2ks})=\prod_p (1-p^{-k(2s-1)})^{-1},$$

which means that $f(s)$ can be obtained by substituting $2s-1$ for s in the factorization of Riemann's $\zeta(s)$. And the residue of $f(s)=\zeta(2s-1)$ at $s=1$ is distinct from 1, since $(s-1)\zeta(s)\rightarrow 1$ and $((2s-1)-1)/(s-1)\rightarrow 2\neq 1$ as $s\rightarrow 1$.

It should be mentioned that a similar example can be derived from a consideration of Hardy [2] concerning the discontinuity of certain expressions involving Ramanujan sums.

6. The theorem italicized in the Introduction will now be deduced from (ii).

Dedekind's zeta-function of an algebraic number field $\mathfrak{K}$ is defined, if $\sigma>1$, by

$$\zeta(s;\mathfrak{K})=\prod (1-(N\mathfrak{p})^{-s})^{-1},$$

where $\mathfrak{p}$ ranges through all prime ideals of $\mathfrak{K}$ and $N\mathfrak{p}$ is the norm of $\mathfrak{p}$. If the factors occurring in this product are arranged in the order corresponding to the principal ideals $[p]$, where p is a rational prime, it is seen that, if $\sigma>1$,

$$\zeta(s;\mathfrak{K})=\prod_p (1-p^{-g_1 s})^{-1}\cdots(1-p^{-g_j s})^{-1}, \tag{19}$$

where p runs through the sequence of all rational primes and g, j denote the positive integers defined in (2), (1).

In the definition of the function (1) of the rational prime p, the rôle of the restriction to *distinct* prime ideals is clear from the fact that the number of *all* prime ideals dividing the principal ideal $[p]$ is independent of p. In fact, if

$$l_1=l_1(p)\geqq 1,\cdots,l_j=l_{j(p)}(p)\geqq 1 \tag{20}$$

denote the multiplicities of the j distinct prime ideals which divide $[p]$ and have the respective degrees (2), then the number of all distinct prime ideals dividing $[p]$ is the sum

$$l_1(p)g_1(p)+\cdots+l_{j(p)}(p)g_{j(p)}(p), \tag{21}$$

which is just m for every p, if m denotes the degree of $\mathfrak{K}$. In addition,

$$g_1(p)+\cdots+g_{j(p)}(p)=m \quad \text{for every} \quad p>d^{\frac{1}{2}}, \tag{22}$$

if d is the discriminant of $\mathfrak{K}$. In fact, each of the j multiplicities (20)

occurring in the decomposition (21) of m is 1 unless p^2 divides the discriminant. This follows from the classical lemma according to which the factorization of a principal ideal $[p]$ into prime ideals contains a multiple factor (if and) only if the rational prime p is a divisor of the discriminant.

Thus it is clear that, if (19) is identified with (9), then the absolute value of $F(p^k)$ is less than a constant multiple of $p^{\theta(1-k)}$ for every fixed $\theta = \epsilon > 0$. It follows therefore from (ii) that the product (3), which results by inserting the factors $1 - p^{-1}$ into the case $s = 1$ of the product (19), is convergent, and that its value is the residue of $\zeta(s; \mathfrak{K})$ at $s = 1$. Since this residue is identical with the asymptotic density (Dirichlet-Dedekind), the proof is complete.

7. What concerns (i), it was mentioned in the Introduction that, in the particular case (19) of (9), the convergence of the series (5) may be obtained from Mertens' theorem for $\mathfrak{K}$. The actual content of (i) admits, in the present case, the following interpretation:

With reference to a fixed algebraic number field $\mathfrak{K}$, let a rational prime p be called "normal" if the number of those *distinct prime ideals of first degree* which occur in the factorization of the principal ideal $[p]$ is exactly 1. And let the remaining, or "abnormal," rational primes (if any) be classified as "defective" or "excessive" according as that number is 0 or at least 2 (it can never be greater than m, the degree of $\mathfrak{K}$). For instance, every p is normal if $\mathfrak{K}$ is the rational field. If $m = 2$, there are various senses in which it is meaningful to say that, as $x \to \infty$, there are in the range $1 < p < x$ about as many primes p for which the discriminant of $\mathfrak{K}$ is quadratic residue as primes p for which it is quadratic non-residue (one of the many possible formulations of this principle, which are of varying degree of analytical "depth," was proved by Pólya [5]). If $\mathfrak{K}$, instead of being (real and) quadratic, is an arbitrary algebraic number field, one will expect a corresponding asymptotic *balance* between the two sets of abnormal primes. But what the convergence of the series (5) means is precisely such a balance.

In fact, let $\nu_p = \nu_p(\mathfrak{K})$, where $p = 2, 3, \cdots$, denote Kronecker's index, defined as the number of those distinct prime ideals of first degree which divide the principal ideal $[p]$ (so that ν_p is non-negative and cannot exceed the degree of $\mathfrak{K}$). Then, if (19) is identified with (9), it is clear from the definition of the positive integers (1), (2) that $F(p)$ can be identified with ν_p. Hence, (i) means the convergence of the series $\Sigma(\nu_p - 1)/p$ (which must be arranged in the order of the monotone sequence of all rational primes). But the factor $\nu_p - 1$ which here multiplies the term $1/p$ of the *divergent* series $\Sigma 1/p$ is $-1, 0$ or at least 1 according as p is defective, normal or excessive.

Thus the *convergence* of the series $\Sigma(\nu_p - 1)/p$ means for the average behavior of the factor $\nu_p - 1$ a specific kind of asymptotic oscillation, representing the "balance" in question.

Appendix.

The integral-valued functions (1), (2), (20) of the rational prime p substantially describe the laws of factorization of the integral ideals of an algebraic number field $\mathfrak{K}$. Conversely, these laws of a given $\mathfrak{K}$ determine the functions (1), (2), (20) uniquely. But fields $\mathfrak{K}$ for which the mystery of these laws has been solved are scarce indeed. And the following considerations imply that, due to this purely algebraical mystery, the difficulties of anything like Riemann's hypothesis in case of an arbitrary algebraic number field $\mathfrak{K}$ go much deeper than in case of a $\mathfrak{K}$ which is quadratic or cyclotomic (or, for that matter, rational).

Let a *degree-function* mean an arbitrary assignment, for every rational prime p, of a positive integer (1) and of $j = j(p)$ positive integers (2), subject to the restriction that the resulting functions j, g of p satisfy (22) for some $m = \text{const.}$, $d = \text{Const.}$

Clearly, every degree-function determines an infinite product (19) which, by uniform convergence, represents a regular function in the half-plane $\sigma > 1$. However, a degree-function is not in general such as to make the product (19) identical with the zeta-function of some algebraic number field $\mathfrak{K}$. In fact, the characterization of these "algebraic" degree-functions presents an arithmetical existence problem, even the nature of which is quite obscure today.

All that is clear is that the arithmetical restrictions in question impose function-theoretical limitations on the choice of the degree-function. For instance, a degree-function cannot belong to an algebraic number field unless the function represented by the product (19) in the half-plane $\sigma > 1$ admits of an analytic continuation regular in the whole plane except for a simple pole at $s = 1$. A further necessary condition is that the resulting meromorphic function be such as to satisfy a Riemann-Hecke functional equation. (Incidentally, it is not evident in itself that this second necessary condition is independent of the first.) However, the existence of a $\mathfrak{K}$ remains undecided even if a functional equation *is* satisfied.

What actually happens is that, *if the degree-function is chosen at random, then the corresponding Eulerian product* (19), *instead of defining a function meromorphic* (*or, at least, algebroid*) *in the whole s-plane, will possess a natural boundary; in fact, this will be the case for "almost all" choices of the degree-function.* The "almost all" refers to a Lebesgue measure naturally associated

with the space of all degree-functions belonging to any fixed value of m in (22).

In fact, consider an arbitrary disjunction of the sequence of all rational primes p into two complementary subsequences, and let a p be called a q or an r according as p is in the first or in the second subsequence. Thus

$$\prod_q (1-q^{-s})^{-1} \prod_r (1-r^{-s})^{-1} = \prod_p (1-p^{-s})^{-1}$$

is Riemann's $\zeta(s)$ and both products

$$\eta(s) = \prod_q (1-q^{-s})^{-1}, \qquad \zeta(s)/\eta(s) = \prod_r (1-r^{-s})^{-1} \tag{23}$$

represent non-vanishing regular functions in the half-plane $\sigma > 1$. It was shown in [3], p. 23, that the underlying disjunction can be so chosen as to make *the line* $\sigma = 1$ *a natural boundary* for the first, and therefore for the second, of the functions (23).

However, the "generic" case, characterizing "almost all" disjunctions, is that the first, and therefore the second, of the corresponding functions (23) remains regular on the line $\sigma = 1$ except for *an algebraic branch-point of order* $\frac{1}{2}$ *(instead of a pole of order* 1) *at* $s = 1$, and admits across the line $\sigma = 1$ an analytic continuation which, except for possible algebraic singularities (the non-existence of which depends on the truth of Riemann's hypothesis for the $\zeta(s)$ of the rational field), exists and is regular in the half-plane $\sigma > \frac{1}{2}$ but has *the line* $\sigma = \frac{1}{2}$ *as natural boundary*; cf. [6], p. 27.

In order to deduce from these facts the assertion italicized before (23), let, with reference to any *given* disjunction, a degree-function belonging to $m = 2$ in (22) be defined as follows: The value of the function (1) of p is chosen to be 2 or 1 according as p is a q or an r. This implies a unique assignment of the functions (2) also, if (22) is required to hold for $m = 2$ (and for every p). It is clear from the identity

$$\prod_q (1+q^{-s})^{-1} \prod_q (1-q^{-s})^{-1} = \prod_q (1-q^{-2s})^{-1},$$

where $\sigma > 1$, that the product (19) defined by the resulting degree-function can be written in the form

$$\prod_q (1-q^{-s})^{-1} \prod_r (1-r^{-s})^{-1} \prod_q (1-q^{-2s})^{-1}$$

and is therefore identical with

$$\zeta(s)\eta(2s),$$

by (23) and by the identity preceding (23).

This completes the proof, since $\zeta(s)$ is meromorphic in the whole plane but $\eta(2s)$ has sometimes the line $\sigma = \frac{1}{2}$, and almost always the line $\sigma = \frac{1}{4}$, as natural boundary.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] R. Fricke, *Lehrbuch der Algebra*, vol. 3, *Algebraische Zahlen*, Braunschweig, 1928.
[2] G. H. Hardy, "Note on Ramanujan's trigonometrical sum $c_q(n)$," *Proceedings of the Cambridge Philosophical Society*, vol. 20 (1921), pp. 263-271.
[3] A. E. Ingham, *The distribution of prime numbers*, Cambridge Tracts, No. 30 (1932).
[4] J. Karamata, "Weiterführung der N. Wienerschen Methode," *Mathematische Zeitschrift*, vol. 38 (1934), pp. 701-708.
[5] G. Pólya, "Ueber die Verteilung der quadratischen Reste und Nichtreste, *Göttinger Nachrichten*, 1918, pp. 21-29.
[6] A. Wintner, *The theory of measure in arithmetical semi-groups*, Baltimore, 1944.
[7] ———, "The densities of the ideal classes and the existence of unities in algebraic number fields," *American Journal of Mathematics*, vol. 67 (1945), pp. 235-238.

THE FUNDAMENTAL LEMMA IN DIRICHLET'S THEORY OF THE ARITHMETICAL PROGRESSIONS.*

By Aurel Wintner.

1. By a completely multiplicative function $\chi = \chi(n)$ is meant any representation of the ordinary multiplication on the semi-group of all positive integers n, that is, any infinite sequence of real or complex numbers $\chi(1)$, $\chi(2), \cdots$ satisfying the identity $\chi(n)\chi(m) = \chi(nm)$ and the additional restriction that $\chi(n) \neq 0$ holds for at least one n. Since the latter proviso is equivalent to $\chi(1) \neq 0$ and also to $\chi(1) = 1$, it follows that the most general $\chi(n)$ results by choosing its values, $\chi(p)$, for primes, p, in an arbitrary manner and then placing

$$\chi(n) = \prod_{p|n} \chi^k(p) \quad \text{if} \quad n = \prod_{p|n} p^k, \tag{1}$$

where $\chi^k(p)$ denotes the k-th power of $\chi(p)$, and both products have the value 1 when they are vacuous, that is, when $n = 1$.

Clearly, (1) is *formally* equivalent to the truth of Euler's relation

$$\sum_{n=1}^{\infty} \chi(n)/n^s = \prod_p (1 - \chi(p)/p^s)^{-1} \tag{2}$$

(as an identity in s). The italicized reservation is necessary, since the series (2) and/or the product (2) can diverge for every s. It may be mentioned that this will not be the case if and only if the values $\chi(p)$ determining the representation $\chi(n)$ are so chosen that there exists a sufficiently large C satisfying

$$|\chi(p)| < p^C \tag{3}$$

for every p. Then (2) holds, by absolute convergence, in the half-plane $\sigma > C + 1$, where $s = \sigma + it$.

A classical instance of (1) is supplied by any of the $\phi(m)$ residue characters (mod m). Any of these particular functions χ of n has the period m and, except in case of a principal character, the mean-value 0 over a period. This implies that $L(s)$, the Dirichlet series on the left of (2), is convergent in the half-plane $\sigma > 0$. In addition, all non-vanishing values of χ become of absolute value 1, that is,

* Received December 10, 1945.

$$|\chi(p)|^2 = |\chi(p)| \tag{4}$$

holds for every p.

If $\chi(n)$ is a complex character $(\mathrm{mod}\, m)$, the non-vanishing of $L(1)$ is trivial; cf., e. g., [1], p. 171. On the other hand, if $\chi(n)$ is a (non-principal) real character (which, by (4), means that $\chi(p)$ is capable of the three values

$$\chi(p) = -1, 0, 1 \tag{5}$$

only, where -1 actually occurs), then the non-vanishing of $L(1)$ is the central fact in Dirichlet's proof for the existence of an infinity of primes in each of the $\phi(m)$ arithmetical progressions h, $h+m$, $h+2m, \cdots$, where $(h, m) = 1$. Although it is undoubtedly impossible *to conclude from Dirichlet's theorem on the progressions that none of the real non-principal values* $L(1)$ *vanishes, (which alone would prove the necessity of an analytical proof)*, the theorem on the progressions has never been reached by an approach distinct from Dirichlet's route. In fact, all the known variants differ only in variations of the proof of the fundamental lemma, $L(1) \neq 0$, of Dirichlet's theory.

2. A short account of these variants may be read on pp. 169-172 in [1]. The literature quoted by Hecke can be completed by referring to a more recent device of Ingham [2], further developed by Rankin [4]. Ingham's method, adjusted to complex-valued representations $\chi(n)$ satisfying (4), consists in taking suitable combinations of complex-conjugates and then applying the functional-theoretical argument used by Landau [3] in the case (5). This elaborate approach has far-reaching advantages (for instance, it leads to a variant of the Hadamard–de la Vallée Pousin proof of $\zeta(1+it) \neq 0$ for $-\infty < t < \infty$). But in the case of real-valued representations $\chi(n)$, it does not seem to apply to cases distinct from (5), since it depends on (4), that is, on the assumption that all non-vanishing values of the (real or complex) function $\chi(n)$ be of absolute value 1.

However, it is suggested by a recent consideration ([5], p. 68, where $f'(n)$ corresponds to the present $\chi(n)$, and $f(n)$ to the coefficient of $1/n^s$ in the Dirichlet series of $\zeta(s)L(s)$; cf. (18) below), that the true condition belonging to a real representation $\chi(n)$ has nothing to do with the sharp *arithmetical* restriction (5), but merely with the assumption

$$-1 \leqq \chi(p) < \infty \tag{6}$$

(for every p), which is *qualitative* in nature. Since this unilateral restriction does not imply the convergence of the Dirichlet series $L(s)$ for some s, an additional convergence condition must then of course be required. But this additional restriction disappears if (6) is particularized to the generalization

$$-1 \leqq \chi(p) \leqq 1 \tag{7}$$

of the classical case (5). And the case (7) is of fundamental *arithmetical* interest, since, as seen from (1), a real-valued representation $\chi(n)$ of the multiplication is a *bounded* representation if and only if its data $\chi(p)$ *satisfy* (7) *for every* p.

It turns out that the extension of Dirichlet's theorem $L(1) \neq 0$ from the classical case (5) to the case (7), and even for the more general case of (6), is actually possible. The proof will depend on an observation the rough content of which is to the effect that *the rôles of the L-series* (2) *and of the Riemann zeta-function*

$$\zeta(s) = \prod_p (1 - 1/p^s)^{-1} \qquad (\sigma > 1) \tag{8}$$

can be interchanged in the Landau-Ingham treatment of the case (4). Correspondingly, what is essential in the analogous situation considered in [5], p. 68, is not the sharp restriction (5), but merely the assumption that the coefficient $1 + a_p = 1 + \chi(p)$ on the right of the formal identity

$$(1 + a_p/p^s)(1 - 1/p^s)^{-1} = 1 + \sum_{k=1}^{\infty} (1 + a_p)/p^{ks}, \tag{9}$$

where $a_p = \chi(p)$ and $\sigma > 0$, be non-negative for every p. But this is precisely the assumption (6).

3. It will be convenient to defer the treatment of the general case (6), which is complicated by the necessity of an additional convergence assumption but is otherwise not different from the treatment of the particular case (7). In the latter case, the theorem is as follows:

(*) *Let* $\chi(n)$ *be a real, bounded representation* (1) *of the ordinary multiplication on the semi-group of the positive integers. Suppose that the regular function* $L(s)$ *defined by the Dirichlet series*

$$L(s) = \sum_{n=1}^{\infty} \chi(n)/n^s \tag{10}$$

on the half-line $s > 1$ *admits across the point* $s = 1$ *an analytic continuation which remains regular on the segment*

$$\tfrac{1}{2} \leqq s \leqq 1. \tag{11}$$

Then

$$L(1) > 0 \tag{12}$$

or, what amounts to the same thing, $L(1) \neq 0$.

The regularity of $L(s)$ *on* (11) *is essential, since* $L(1)=0$ *becomes possible as soon as* (11) *is replaced by*

$$\tfrac{1}{2} < s \leqq 1. \tag{13}$$

The insufficiency of (13) is proved by the standard case of Liouville's function, $\chi(n)=\lambda(n)$. In this case, (1) is satisfied, as is the boundedness condition (7), since $\lambda(p)=-1$ for every p (incidentally, this is just *the extreme case admitted by* (6) *alone*). Hence, if s is replaced by $2s$ in (8), it is seen from (2) that the function (10), where $s>1$, now becomes

$$L(s)=\prod_p (1+1/p^s)=\zeta(2s)/\zeta(s)$$

and is therefore regular on the segment (13). Nevertheless, $L(1)=0$, since $\zeta(s)$ in $\zeta(2s)/\zeta(s)$ has a pole at $s=1$ but $\zeta(2)\neq 0$.

In the sequel, only (11) will be considered.

First, the boundedness of $\chi(n)$, which is equivalent to (7), implies that both the Dirichlet series (10) and its Eulerian factorization (2) are absolutely convergent in the half-plane $\sigma>1$. Furthermore, it is clear from (7), and even from the more general assumption (6), that every factor of the product (2) is positive when $s>1$ (simply because $1/p^s<1$). Since the function $L(s)$, being regular for $s\geqq\frac{1}{2}$, is continuous at $s=1$, it follows that the assertion (12) is equivalent to negation of

$$L(1)=0. \tag{14}$$

4. It is also clear from (7) and (1) that the Dirichlet series

$$J(s)=\sum_{n=1}^{\infty}\chi^2(n)/n^s, \tag{15}$$

where $\chi^2(n)$ denotes the square of $\chi(n)$, is (absolutely) convergent on the half-line $s>1$ and admits there the Eulerian factorization

$$J(s)=\prod_p (1-\chi^2(p)/p^s)^{-1}. \tag{16}$$

If s in (16) is replaced by $2s$, it is seen from the factorization (2) of $L(s)$ that

$$L(s)/J(2s)=\prod_p (1+\chi(p)/p^s)$$

when $s>1$. Hence, if $\beta(n)$ is defined by

$$\zeta(s)L(s)/J(2s)=\sum_{n=1}^{\infty}\beta(n)/n^s, \tag{17}$$

then, from (8),

$$(18) \qquad \prod_p (1+\chi(p)/p^s)(1-1/p^s)^{-1} = \sum_{n=1}^{\infty} \beta(n)/n^s.$$

Finally, if the case $a_p = \chi(p)$ of the identity (9) is combined with the assumption (7), it is seen from (18) that

$$(19) \qquad \beta(n) \geqq 0$$

holds for every n and, of course,

$$(20) \qquad \beta(1) = 1.$$

Suppose that (14) is true.

Since the function $L(s)$ is supposed to be regular on the segment (12), and therefore on the half-line $s \geqq \frac{1}{2}$, and since the zero of $L(s)$ at $s = 1$ absorbs the pole of $\zeta(s)$ in the product $\zeta(s)L(s)$, it follows that this product is regular for $s \geqq \frac{1}{2}$. On the other hand, since (16) represents a non-vanishing regular function on the half-line $s > 1$, the function $1/J(2s)$ is regular for $s > \frac{1}{2}$. Consequently, the function on the left of (17) is regular for $s > \frac{1}{2}$. It follows therefore from (19) and from Landau's extension [3] of the Vivanti-Pringsheim theorem to Dirichlet series, that the Dirichlet series on the right of (17) is convergent for $s > \frac{1}{2}$. And, for reasons of analyticity, the sum of this Dirichlet series must be the function on the left of (17), if $s > \frac{1}{2}$.

It turns out that this contains a contradiction. In order to show this, the alternative possibilities represented by

$$(21) \qquad \sum_p \chi^2(p)/p < \infty$$

and

$$(22) \qquad \sum_p \chi^2(p)/p = \infty$$

will have to be disposed of separately.

5. Consider first the possibility (21). In this case, it is clear that the sum of the reciprocal values of those primes p which satisfy the inequality $|\chi(p)| \geqq \frac{1}{2}$ is a convergent series, since each of the corresponding terms of (21) is minorized by the reciprocal value of $4p$. But it will now be shown that (whether (21) or (22) be the case) the sum of the reciprocal values of those primes p which satisfy the complementary inequality, $|\chi(p)| < \frac{1}{2}$, must also

converge. Hence, (21) will lead to the contradiction $\Sigma 1/p < \infty$, where p runs through all primes.

First, since the Dirichlet series on the right of (17) is convergent for $s > \frac{1}{2}$, and therefore for $s = 1$, it follows from (19) that this Dirichlet series is absolutely convergent for $s = 1$. On the other hand, since (15) is absolutely convergent for $s > 1$ and therefore for $s = 2$, the Dirichlet series of $J(2s)$ is absolutely convergent at $s = 1$. It follows therefore from (17) that $\zeta(s)L(s)$ is the product of two ordinary Dirichlet series both of which are absolutely convergent at $s = 1$. Hence, the same is true of the product series. In particular, if c_n denotes the coefficient of $1/n^s$ in the Dirichlet series of $\zeta(s)L(s)$, then $\Sigma' | c_p |/p < \infty$, where p runs through all primes. But it is clear from (8) and from the factorization (2) of $L(s)$, where $s > 1$, that $c_p = 1 + \chi(p)$. Accordingly,

$$\sum_p | 1 + \chi(p) |/p < \infty.$$

However, this implies that the sum of the reciprocal values of those primes p which satisfy the inequality $|\chi(p)| < \frac{1}{2}$ is a convergent series. In fact, $| 1 + \chi(p) |$ is then minorized by $\frac{1}{2}$.

This completes the proof for the impossibility of (21). The refutation of the remaining possibility, (22), is quite different. It proceeds as follows:

Since the Dirichlet series on the right of (17) converges to the function on the left of (17) if $s > \frac{1}{2}$, it is clear from (19) and (20) that

$$\zeta(s)L(s) \geqq J(2s)$$

if $s > \frac{1}{2}$. Since (16) is valid for $s > 1$, it follows that

$$\zeta(s)L(s) \geqq 1/\prod_p (1 - \chi^2(p)/p^{2s})$$

if $s > \frac{1}{2}$. Hence it is seen from $\chi^2(p) \geqq 0$ that (22) implies the relation

$$\liminf_{s \to \frac{1}{2}+0} \zeta(s)L(s) = \infty.$$

But this relation contradicts the assumption (which has not been used thus far) that the function $L(s)$ is regular not only on the segment (13) but on the closed segment (12) as well.

This completes the proof of (*).

Remark. *The assertion of* (*) *remains valid if the function* $L(s)$,

instead of being regular on the segment (12), *is regular just on the segment* (13), *but does not tend to* ∞ *as* $s \to \frac{1}{2} + 0$ (*which does not even require that* $L(s) <$ const. *as* $s \to \frac{1}{2} + 0$).

In fact, all that is needed is that the assertion of the last formula line be contradictory.

6. It is now easy to see that the assumption (7) of (i) can be generalized to (6), provided that a convergence condition is satisfied.

(**) *Let a representation* (1) *of the ordinary multiplication on the semi-group of all positive integers* n *be such as to satisfy the unilateral restriction* $\chi(p) \geqq -1$ *for every prime* p. *Suppose that the Dirichlet series*

$$J(s) = \sum_{n=1}^{\infty} \chi^2(n)/n^s,$$

where $\chi^2(n)$ *denotes the square of* $\chi(n)$, *is convergent for* $s > 1$. *Then the same is true of the Dirichlet series*

$$L(s) = \sum_{n=1}^{\infty} \chi(n)/n^s.$$

Suppose that the function $L(s)$ *admits across the point* $s = 1$ *an analytic continuation which remains regular for* $\frac{1}{2} \leqq s \leqq 1$. *Then* $L(1) \neq 0$ (*or, equivalently,* $L(1) > 0$).

In fact, the convergence of the Dirichlet series $J(s)$ for $s > 1$ implies, by Schwarz's inequality, the absolute convergence of the Dirichlet series $L(s)$ for $s > 1$. Hence, the Eulerian factorizations (2), (16) hold, by absolute convergence, for $s > 1$. Since $\chi(p) \geqq -1$, none of the factors occurring in these factorizations of $L(s), J(s)$ becomes meaningless (i. e., of the form 0^{-1}). Correspondingly, the proof of (**) is exactly the same as was that of (*). In fact, the various steps in the proof of (*) were purposely referring to (6) rather than to (7).

Remark. The whole proof is such as to leave little doubt that the value -1 of the absolute constant occurring on the right of the assumption $\chi(p) \geqq -1$ of (ii) is as sharp as possible. That *some* lower limitation $\chi(p) \geqq -\theta$ is necessary, is shown by the example $\chi(n)$ defined by $\chi(2) = -2$, $\chi(3) = \chi(5) = \cdots = 0$. In fact, (14) is true in this case, since the factori-

zation (2) shows that $L(s)$ becomes identical with the function $1-2^{1-s}$. Thus $\theta=2$ is not an admissible value of θ in $\chi(p)\geqq-\theta$. But all that follows in this trivial manner does not disprove the (unlikely) possibility $\theta>1$. In this regard, constructions of the type considered in [6] seem to be relevant.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] E. Hecke, *Vorlesungen über die Theorie der algebraischen Zahlen*, Leipzig, 1923.

[2] A. E. Ingham, "Note on Riemann's zeta-function and Dirichlet's L-functions," *Journal of the London Mathematical Society*, vol. 5 (1930), pp. 107-112.

[3] E. Landau, "Über einen Satz von Tschebyschef," *Mathematische Annalen*, vol. 61 (1905), pp. 527-550.

[4] R. A. Rankin, "Contributions to the theory of Ramanujan's tau-function and similar arithmetical functions, I," *Proceedings of the Cambridge Philosophical Society*, vol. 35 (1939), pp. 351-356.

[5] A. Wintner, *Eratosthenian Averages*, Baltimore, 1943.

[6] ———, "The singularities in a family of zeta-functions," *Duke Mathematical Journal*, vol. 11 (1944), pp. 287-291.

ASYMPTOTIC INTEGRATION CONSTANTS IN THE SINGULARITY OF BRIOT-BOUQUET.*

By Aurel Wintner.

1. The following considerations deal with the problem of the real Briot-Bouquet equation

$$xy' = py + qx + \phi(x, y), \tag{1}$$

where the prime denotes differentiation with respect to x, the coefficients p, q are constants and $\phi(x, y)$ represents terms which, in a sense to be specified, are of a higher order than the linear form $py + qx$, as $(x, y) \to (0, 0)$. The connection with Poincaré's problem

$$\dot{x} = ax + by + f(x, y), \qquad \dot{y} = cx + dy + g(x, y), \tag{2}$$

where the dots denote differentiations with respect to a time variable t $(\to \infty)$, the constant matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{3}$$

has a non-vanishing determinant and $f(x, y)$, $g(x, y)$ represent the "small" non-linear perturbations, is as follows:

The geometrical problem concerning the behavior of the solution paths of (2) near $(x, y) = (0, 0)$ naturally splits into two main types, according as the characteristic numbers of (3) are or are not real *and* of the same sign. The second type comprises three subcases, since the characteristic numbers can then be real and of opposite signs (saddle), complex but not purely imaginary (vortex) or purely imaginary (center *or* vortex). The problem to be treated does not arise in any of these subcases of the second type, treated, under very general assumptions, in the second part of Perron's paper [4] (correspondingly, all of the following references to [4] will quote its first part only). In fact, the problem to be considered is one concerning a node, represented by the first type, that is, by the cases in which the characteristic numbers of (3) are real and of the same sign. This type, too, comprises three subcases, since the characteristic numbers can now be distinct or not and, if they are not distinct, the elementary divisors can be distinct or not (the corresponding normal forms of (3) are

* Received December 14, 1945.

$$(4) \qquad \begin{pmatrix} 1 & 0 \\ 0 & l \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix},$$

where $0 < l \neq 1$). And the connection referred to before (2) consists in the formal fact that (2) can be reduced, in all three nodal cases (4), to (1), where $p > 0$ (and, without substantial loss of generality,

$$(4\text{ bis}) \qquad p = 1,\ q = 0; \quad p > 1,\ q = 0; \quad p = 1,\ q = 1$$

respectively); cf., e. g., [4], pp. 140-146.

The six figures, illustrating the three possibilities in the three subcases (saddle, vortex, center) of the second type and the three distinct kinds of nodes of the first type, may be found in [2], pp. 101-102. The three nodal figures are given also in [4], p. 123. For references to the classical literature of the subject, cf. [3], pp. 215-216, 219-220, 227-228. The more recent papers of Frommer [1] and Weyl [5] do not deal with the problem to be considered.

2. The problem in question concerns *the determination of the paths in a nodal sheaf by the asymptotic slope (or its equivalent) as an arbitrary integration constant.* In all three cases (4 bis), this problem will be treated as an application of a general theorem on asymptotic equilibria (cf. [6]), including a continuity theorem of Siegel (ibid., p. 131) concerning such equilibria. When formulated for the relevant case of a single differential equation, these facts can be stated as follows:

In a half-plane

$$(5) \qquad u_0 < u < \infty, \qquad -\infty < v < \infty,$$

let $f(u, v)$ be a real-valued, continuous function satisfying

$$(6) \qquad \int^{\infty} |f(u, 0)|\, du < \infty$$

and having the property that, if u is large enough and v' and v'' are arbitrary, the Lipschitz condition

$$(7) \qquad |f(u, v') - f(u, v'')| \leqq \lambda(u)\, |v' - v''|$$

is satisfied by a (non-negative and, for instance, continuous) function λ for which

$$(8) \qquad \int^{\infty} \lambda(u)\, du < \infty.$$

Then, if v^0 is given arbitrarily and if $u^0(> u_0)$ is any value greater than a lower bound depending on the given value of v^0, the differential equation

(9) $$dv/du = f(u, v)$$

and the initial condition $v(u^0) = v^0$ determine a solution $v = v(u)$ which exists on the whole half-line $u^0 < u < \infty$. Furthermore, there exists a finite limit, $v(\infty)$, for this solution. Conversely, if c is any real number, there exists a solution $v = v(u)$ corresponding to which the limit $v(\infty)$ attains the given value c, and this solution is uniquely determined by c. In addition, the correspondence between c and the solution $v = v(u)$ determined by $v(\infty) = c$ is continuous.

[A corresponding theorem holds for the case in which (9) is replaced by a system, that is, in which v, f are vectors. Theorem (ii) in [6] extends this to the case in which the asymptotic equilibria, $v(\infty) = c$, are replaced by small oscillations about such equilibria. This extension depended on the method of the variation of constants in the particular case of small oscillations. Correspondingly, a further extension results if, for the case of inhomogeneous *linear* systems, the method of the variation of constants is formulated in its general form, as follows: Let $A = A(t)$ be a matrix of n times n continuous functions on a t-interval and let $a(t)$ be a vector of n continuous functions on the same interval. Then, if $X = X(t)$ is any fixed fundamental matrix of the homogeneous linear system $x' = A(t)x$ (where $x = (x_1, \cdots, x_n)$ and $x' = dx/dt$), that is, if $X(t)$ is a matrix the n columns of which are n linearly independent solution vectors $x(t)$ of $x' = A(t)x$, then the general solution of the inhomogeneous linear system $y' = A(t)y + a(t)$ is the vector

$$y(t) = X(t)(\alpha + \int_{t_0} X^{-1}(t)a(t)\,dt),$$

where α is an arbitrary constant vector, t_0 is any fixed point of the t-interval (it is understood that the reciprocal matrix, $X^{-1}(t)$, exists for all t, since $\det X(t) \neq 0$ for all t). The verification requires nothing but a differentiation, since $X'(t) = A(t)X(t)$ is an identity in virtue of $x'(t) = A(t)x(t)$.]

3. The prototype of (1) is the differential equation in which the higher terms, represented by $\phi(x, y)$, are missing; so that

(1 bis) $$xy' = y; \qquad xy' = py \quad (p > 1); \qquad xy' = y + x$$

in the respective three cases (4 bis).

In the first case, $xy' = y$, the general solution is $y(x) = cx$, where the

constant c is arbitrary. For the corresponding perturbed equation (1), the following theorem will be proved:

(i) *On a rectangle*

$$0 < x < a, \quad -b < y < b, \tag{10}$$

let $\phi(x, y)$ *be a real-valued, continuous function, subject to the restriction*

$$\int_{+0} x^{-2} \, | \phi(x, 0) | \, dx < \infty \tag{11}$$

and, for sufficiently small $x > 0$, *to the Lipschitz condition*

$$| \phi(x, y_1) - \phi(x, y_2) | \leqq \mu(x) \, | y_1 - y_2 |, \tag{12}$$

where $\mu(x)$ *is defined (and, for instance, continuous) for small* $x > 0$ *and satisfies*

$$\int_{+0} x^{-1} \mu(x) \, dx < \infty. \tag{13}$$

Then, if a and b in (10) *are chosen small enough, the behavior of the solutions of*

$$xy' = y + \phi(x, y) \tag{14}$$

can be described as follows: Every solution path $y = y(x)$ *of* (14), *issuing from an arbitrary point* (x_0, y_0) *of the rectangle* (10), *exists on the whole interval* $0 < x < x_0$ *and tends to the origin,* $(0, 0)$, *in such a way that there exists a constant* $c = c(x_0, y_0)$ *satisfying*

$$y(x) \sim cx \quad \text{as} \quad x \to +0 \tag{15}$$

(*which should mean* $y(x) = o(x)$ *if* $c = 0$). *Conversely, if the (real) value of the integration constant c is assigned arbitrarily, there belongs to it a unique solution path* $y = y(x)$ *satisfying* (15).

It is instructive to contrast this theorem with the corresponding result of Perron (Satz 4 in [4], p. 132). He assumes that $\phi(x, y)$ is continuous on the *closure* of the open rectangle (10), which of course is a serious restriction of the character of the singular point, $(0, 0)$. On the other hand, instead of assuming a Lipschitz condition (12) and the *average* restrictions (13) and (11), Perron requires the existence of a constant $\theta > 0$ satisfying

$$\phi(x, y) = o(r^{1+\theta}) \quad \text{as} \quad r = (x^2 + y^2)^{\frac{1}{2}} \to 0. \tag{16}$$

Since (15) is a condition *at* the origin $(0, 0)$, it does not imply anything like a Lipschitz condition. Correspondingly, it cannot ensure that (14) has just one solution path through a point of the open rectangle (10). Accordingly, the assertion of Perron's theorem is that (14) has at least one (rather than, as in (i), exactly one) solution path satisfying (15), if c is given arbitrarily. But the *explicit* assumption of a fixed $\theta > 0$ in (16), no matter how drastic, seems to be essential in the proof of Perron's theorem. On the other hand, the *integral* assumptions of (i) prove to be the best possible restrictions of their kind. This can be seen as follows:

Suppose that $\phi(x; y)$ in (14) is independent of y, that is, let

$$xy' = y + \phi(x), \tag{17}$$

where $\phi(x)$ is continuous for small $x > 0$ (as will be seen in a moment, it is, this time, beside the main point that, so far as (i) is concerned, $x > 0$ could be replaced by $x \geqq 0$). Since (12) and (13) are now satisfied (by $\mu \equiv 0$), the assumptions of (i) reduce to (11) alone, that is, to the absolute convergence of the improper integral

$$\int_{+0} x^{-2}\phi(x)\,dx. \tag{18}$$

In contrast, Perron's assumption (16) requires that (18) be convergent for the drastic reason that the function integrated in (18) becomes $o(x^{-\alpha})$ for some $\alpha < 1$. On the other hand, if $\phi(x) > 0$, the sufficient condition supplied by (i) turns out to be necessary for the truth of the assertions of (i). In fact, the absolute convergence of (18) is equivalent to the convergence of (18), if $\phi(x) > 0$. But the convergence of (18) is now necessary and sufficient for the truth of the assertions of (i), whether $\phi(x) > 0$ be satisfied or not. This assertion, which proves that (i) is of a final nature, can be verified as follows:

Since the differential equation (17) is linear, it can be solved by a quadrature. This gives

$$y(x) = x\left(c + \int_{x_0} x^{-2}\phi(x)\,dx\right), \tag{18 bis}$$

where x_0 is fixed and c is arbitrary. But the representation (18 bis) of the general solution of (17) makes it clear that the convergence of (18) is necessary and sufficient for the truth of the assertion (15), where c is unspecified (in fact, the parenthetical remark following (15) requires the inclusion of $c = 0$).

Incidentally, all of this implies that, as verified by Perron ([4], p. 131)

by considering a specific $\phi(x)$ in (17), his assumption (16), where $1+\theta>1$, cannot be relaxed to $\phi(x,y)=o(r)$, where $\theta=0$.

4. In the third of the cases (4 bis), the prototype ($\phi\equiv 0$) of (1) is $xy'=y+x$. If $x>0$, the general solution of this linear differential equation is seen to be $y(x)=x\log x+cx$, where c is arbitrary. The appearance of the logarithmic (or, on the standard exponential scale of dynamics, "secular") term agrees, of course, with the fact that the present case, being represented by the third of the matrices (4), is the case of a multiple elementary divisor.

Corresponding to this form of the general solution of the linear prototype, the non-linear situation now turns out to be as follows:

(ii) *Suppose that* $\phi(x,y)$ *satisfies the assumptions of* (i), *except that* (11) *is replaced by*

$$\int_{+0} x^{-2}\,|\,\phi(x,x\log x)\,|\,dx<\infty. \tag{19}$$

Then the assertions of (i) *remain true if*

$$xy'=y+x+\phi(x,y) \tag{20}$$

and

$$y(x)=x\log x+cx+o(x)\quad\text{as}\quad x\to+0 \tag{21}$$

are read instead of (14) *and* (15) *respectively.*

If (20) is chosen to be of the particular type

$$xy'=y+x+\phi(x), \tag{22}$$

the linear differential equation which now replaces (17), then the quadrature which corresponds to the solution (18 bis) of (17) in the case (22) proves that (ii) is of a final nature in the same sense as (i).

What concerns Perron's result in this case of a multiple elementary divisor (Satz 5 in [4], p. 133), his assumptions are, as before, the continuity of $\phi(x,y)$ on the *closure* of the rectangle (10) and the existence of a positive index θ satisfying (16). But his assertion is now very weak; in fact, not even the first approximation, $y(x)\sim x\log x$, to (21), but merely the qualitative corollary, $y(x)/x\to-\infty$, of this approximation is asserted. In contrast, (ii) introduces c as an asymptotic integration constant and establishes a one-to-one correspondence between the field of the solution paths and the points of the line $-\infty<c<\infty$.

5. The remaining case is the second in (4 bis). According to (1), the corresponding linear prototype is $xy'=py$, where $p>1$. If $x>0$, the general

solution of this linear differential equation is $y(x) = cx^p$ (whether p does or does not satisfy the assumption $p > 1$, except that

$$y(x) = o(x) \quad \text{as} \quad x \to +0, \quad \text{if} \quad p > 1, \tag{23}$$

but not if $p \leqq 1$). Correspondingly, the non-linear theorem now turns out to be as follows:

(i*) *For every fixed $p > 0$, the assertions of* (i) *remain true if* (11) *is replaced by*

$$\int_{+0} x^{-1-p} \, | \phi(x, 0) | \, dx < \infty, \tag{11*}$$

and (14) *by*

$$xy' = py + \phi(x, y), \tag{14*}$$

finally (15) *by*

$$y(x) = cx^p + o(x^p) \quad \text{as} \quad x \to +0. \tag{15*}$$

This is a generalization of (i), since $p = 1$ (and, for that matter, $0 < p < 1$) is allowed in (i*). If $p > 1$, then (23) is a corollary of the assertion of the last formula line.

Perron's corresponding result (Satz 3 in [4], p. 129) concerns the case $p > 1$. Besides the continuity of $\phi(x, y)$ on the *closure* of (10), all that he now assumes is

$$\phi(x, y) = o(r) \quad \text{as} \quad r = (x^2 + y^2)^{\frac{1}{2}} \to 0 \tag{16 bis}$$

(that is, (16) in the limiting case $\theta = 0$, excluded in the preceding cases), but his assertion is just (23), a relation containing no integration constant.

6. Since (i*) implies (i), it will suffice to prove (i*) and (ii).

In both proofs, it can be assumed that b is ∞ in (10). In fact, if $\phi(x, y)$ is given as a continuous function on a rectangle (10), it can be assumed to be extended to the corresponding strip

$$0 < x < a, \qquad -\infty < y < \infty \tag{24}$$

in such a way that the respective assumptions of (i*) and (ii) remain satisfied. But, if the assertions of (i*) and (ii) are proved for the case in which (10) is replaced by (24), the assertions of (i*) and (ii) themselves follow, since these assertions only concern a sufficiently small rectangle (10).

By the method of the variation of constants, (i*) and (ii) will now be reduced to the theorem quoted after (5). To this end, put

$$x = e^{-u}, \qquad y = ve^{-pu} \tag{25_1}$$

in the differential equation, (14*), of (i*), and

$$(25_2) \qquad x = e^{-u}, \qquad y = (v - u)e^{-u}$$

in the differential equation, (20), of (ii). Both (25_1) and (25_2) transform the strip (24) into a (u, v)-domain which can be assumed to be the (*schlicht*) half-plane (5) of the theorem to be applied. Both differential equations then appear in the form (9), where, as easily verified,

$$(26_1) \qquad f(u, v) = -e^{pu}\phi(e^{-u}, ve^{-pu})$$

in the case of (i*) and

$$(26_2) \qquad f(u, v) = -e^{u}\phi(e^{-u}, (v - u)e^{-u})$$

in the case of (ii).

On the other hand, it is seen from (25_1) and (25_2) that the assertions of (i*) and (ii), those concerning (15*) and (21) respectively, become identical with the corresponding assertions of the theorem quoted in **2**, that is, with those concerning $v(\infty) = c$. Consequently, all that remains to be ascertained is that the assumptions (6), (7), (8) of **2** are satisfied in the respective cases.

What concerns (6), it is clear from (26_1) and (26_2), respectively, that the corresponding assumptions, (11*) and (19), of (i*) and (ii) are identical with (6). On the other hand, straightforward reductions show that the common Lipschitz assumptions, (12) and (13), of (i*) and (ii) are precisely the Lipschitz assumptions, (7) and (8), of **2**, whether $f(u, v)$ be the function (26_1) or the function (26_2).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Frommer, "Die Integralkurven einer gewöhnlichen Differentialgleichung erster Ordnung in der Umgebung rationaler Unbestimmtheitsstellen," *Mathematische Annalen*, vol. 99 (1928), pp. 222-272.

[2] H. Liebmann, *Lehrbuch der Differentialgleichungen*, Leipzig, 1901.

[3] P. Painlevé, "Gewöhnliche Differentialgleichungen; Existenz der Lösungen," *Encyklopädie der mathematischen Wissenschaften*, II A 4 a (1900).

[4] O. Perron, "Ueber die Gestalt der Integralkurven einer Differentialgleichung erster Ordnung in der Umgebung eines singulären Punktes," *Mathematische Zeitschrift*, vol. 15 (1922), pp. 121-146 and vol. 16 (1923), pp. 273-295.

[5] H. Weyl, "Concerning a classical problem in the theory of singular points of ordinary differential equations," *Actas de la Academia Nacional de Ciencias Exactas, Fisicas y Naturales de Lima*, vol. 7 (1944), pp. 21-60.

[6] A. Wintner, "Asymptotic equilibria," *American Journal of Mathematics*, vol. 68 (1946), pp. 125-132.

ON THE ASYMPTOTIC BEHAVIOR OF THE SOLUTIONS OF A NON-LINEAR DIFFERENTIAL EQUATION.*

By Philip Hartman and Aurel Wintner.

1. The following considerations imply a simplified and, at the same time, generalized approach to certain of Poincaré's qualitative results on the singularity $(x, y) = (0, 0)$ of the real (analytic) differential equation

$$xy' = \alpha x + \beta y + \cdots, \qquad (\beta \neq 0), \tag{1}$$

and to the corresponding results of Bendixson on the non-analytic case of this differential equation and on

$$x^m y' = \alpha x + \beta y + \cdots \qquad (\beta \neq 0;\ m = 1, 2, \cdots) \tag{2}$$

(cf. Liebmann's report [2], pp. 507-512, where further references will be found; for more recent literature, cf. p. 178 of Dulac's monograph [1]).

The most general result in the direction in question is that of Perron ([3]; cf. [2], pp. 512-513), which concerns the differential equation

$$\phi(x) y' = f(x, y) \tag{3}$$

in which $\phi(x)$, instead of being x or $x^2, x^3, \cdots$ (as in (1) or (2) respectively), is any function which is continuous and positive on an interval $0 < x \leqq a$ and satisfies the conditions

$$\int_{+0} dx/\phi(x) = \infty \tag{4}$$

and

$$\phi(x) \to 0 \quad \text{as} \quad x \to +0, \tag{5}$$

whereas $f(x, y)$ is a real-valued continuous function on a rectangle

$$0 \leqq x \leqq a, \qquad -b \leqq y \leqq b \tag{6}$$

and is subjected there to the upper *and* lower Lipschitz conditions

$$0 < c < |f(x, y^{**}) - f(x, y^*)| / |y^{**} - y^*| < C, \qquad (y^* < y^{**}), \tag{7}$$

* Received January 8, 1946.

and to the condition

$$f(0,0)=0. \tag{8}$$

Perron distinguishes two cases, according as the sign of the absolute value can or cannot be omitted in (7). In the first case, he proves the existence of two positive numbers a', b' having the property that *every* solution path $y=y(x)$ of (3) issuing from a point of the rectangle

$$R'\colon \quad 0<x<a',\ -b'<y<b', \qquad (a'\leqq a,\ b'\leqq b), \tag{9}$$

tends to the point $(0,0)$ as $x\to+0$. In the second case, he proves that there exists *exactly one* such solution path.

Perron's method of proof consists of an elaborate application of the process of successive approximations. The nature of such an analytical approach makes sufficiently clear the rôle of the upper Lipschitz limitation in (7), that is, of the assumption of a constant C. Correspondingly, this condition is satisfied as soon as $f(x,y)$ is sufficiently smooth (for instance, such as to possess a continuous partial derivative with respect to y). In contrast, the lower Lipschitz in (7), that represented by the existence of a constant $c>0$, is a serious *qualitative* restriction. In fact, it is clear from the continuity of $f(x,y)$ that, if the lower limitation of (7) is satisfied, then $f(x,y)$ is a strictly monotone function of y (for every fixed x), it being increasing in the first, and decreasing in the second, of the cases which represent Perron's alternative, quoted above. But the converse is not true, since, even if $f(x,y)$ is regular-analytic, the monotony of $f(x,y)$ with respect to y does not imply the existence of a constant $c>0$ satisfying (7) near the critical point, $(0,0)$.

2. Thus it appears unexpected that, for the truth of Perron's alternative, *both* the upper and the lower Lipschitz limitations in (7) turn out to be entirely superfluous by virtue of that corollary of (7) which requires the strict monotony of $f(x,y)$ with respect to y. What makes the methodical situation particularly striking is the fact that the proof of the resulting extensions (which apply to new cases even when $f(x,y)$ is restricted to be regular-analytic) can be obtained by a transparent argument of purely *geometrical* considerations. In fact, the proof will require only an extension of the qualitative procedure developed in [4]. A recourse to the *analytical* process of successive approximations proves, therefore, to be a ballast in every respect.

Incidentally, it is superfluous to assume the restriction (5) which, in view of the assumption (4), involves a limitation of the asymptotic smoothness of the coefficient of y' in (3) (the assumption (4), where $\phi(x)>0$, being compatible even with

(5 bis) $$\limsup \phi(x) = \infty$$

as $x \to \infty$). Accordingly, the first case of the alternative to be proved can be formulated as follows:

(i) *Let $\phi(x)$, where $0 < x \leqq a$, be a positive, continuous function satisfying* (4), *and let $f(x, y)$, where (x, y) is restricted to the rectangle* (6), *be a continuous function which satisfies* (8) *and is increasing with y when x is fixed. Then* (6) *contains a rectangle* (9) *(even one with $b' = b$) having the following property: Every solution path $y = y(x)$ of* (3) *passing through a point (x_0, y_0) of* (9) *can be continued for all positive $x (< x_0)$, and*

(10) $$y(x) \to 0 \quad \text{as} \quad x \to +0$$

holds for all continuations.

Such an oblique formulation (in terms of "continuations") is necessary, since, in contrast to the case of an upper Lipschitz limitation on $f(x, y)$, a solution path $y = y(x)$ of (3) can now easily lead to branch points, that is, to points (x_0, y_0) through which there is more than one solution path (suffice it to say that $(u, v) = (0, 0)$ is not a point of uniqueness of the differential equation $dv/du = v^{1/3}$, although $v^{1/3}$ is an increasing function of v).

The second assertion to be proved is as follows:

(ii) *Suppose that $\phi(x)$ and $f(x, y)$ satisfy the assumptions of* (i), *except that $f(x, y)$, instead of being increasing, is decreasing (in y) when x is fixed. Then* (3) *has one and only one solution $y = y(x)$ satisfying* (10).

The fate of all the other solution paths issuing from points of (6) is explained by the following refinement of (ii): If $\phi(x)$ and $f(x, y)$ satisfy the assumptions of (ii) and if $y = y(x)$ is any solution path of (3) distinct from the unique solution path supplied by (ii), then the solution path cannot reach the y-axis within the rectangle (6) and must, therefore, reach one of the levels $y = \pm b$ at a positive $x = x^0$, which is an integration constant.

Under the assumptions of (i) or (ii), the function $f(x, y)$, hence $f(0, y)$ as well, is continuous and strictly monotone (in y) when $-b \leqq y \leqq b$. The assumption (8) implies that $f(0, y)$ is not zero if $y \neq 0$ and so, by continuity, $f(x, y)$ is not zero for sufficiently small x if $y \neq 0$. Furthermore, $f(0, b)$ and $f(0, -b)$ are of opposite signs and, again by virtue of the continuity of $f(x, y)$, it is possible to choose a positive number $a' (\leqq a)$ so small that $f(x, b)$, $f(x, -b)$ do not vanish when $0 \leqq x \leqq a'$ and are of opposite signs.

3. It turns out that these properties of $f(x,y)$ are essentially all that is needed for the truth of the assertions of (i) and (ii). In fact, even the continuity of $f(x,y)$ at any point of the line $x=0$ can be replaced by considerably less restrictive conditions, to the effect that, barring the immediate vicinity of the origin, $f(x,y)$ is bounded away from zero. The fundamental assumptions can then be formulated as follows:

ASSUMPTION (*). *Let $\phi(x)$, where $0<x\leqq a$, be a positive, continuous function satisfying (4). Let $f(x,y)$ be defined and continuous on the (partly open) rectangle*

$$\text{(II)} \qquad R: \quad 0<x\leqq a, \qquad -b\leqq y\leqq b,$$

and have the property that $1/f(x,y)$, outside of any fixed vicinity of the origin, is bounded for sufficiently small x, and that $f(x,b)$ and $f(x,-b)$ do not vanish for $0<x\leqq a$ and are of opposite signs.

Clearly, (i) is contained in the following theorem:

(I) *If $\phi(x)$ and $f(x,y)$ satisfy (*), and if $f(x,b)>0$ and $f(x,-b)<0$ when $0<x\leqq a$, then every solution path $y=y(x)$ of (3) issuing from an interior point (x_0,y_0) of the rectangle (11) can be continued for all positive $x(<x_0)$, and (10) holds for all continuations.*

Correspondingly, the existence statement of (ii) is contained in the following:

(II) *If $\phi(x)$ and $f(x,y)$ satisfy (*) and if $f(x,b)<0$ and $f(x,-b)>0$ when $0<x\leqq a$, then there exists for $0<x\leqq a$ at least one solution path $y=y(x)$ of (3), and (10) holds for all such paths.*

Finally, the uniqueness statement in (ii) is implied by the following:

(II bis) *If $\phi(x)$ and $f(x,y)$ satisfy the conditions of (II) and, in addition, $f(x,y)$ is a non-increasing function of y for every fixed $x>0$, then only one solution path $y=y(x)$ of (3) can satisfy (10).*

4. In order to prove (I) and (II), it will first be shown in either case that *if a solution $y=y(x)$ of (3) exists for all small $x>0$, then (10) is satisfied by it.*

First, if $y=y(x)$ is a solution for which the limit $y(+0)$ exists, then this limit is necessarily zero. For if this were not the case, it would follow that

$1/|f(x, y(x))| < C$ holds for every sufficiently small $x > 0$ and for some $C > 0$. But (3) shows that $y'(x)$ is of constant sign. Hence,

$$\int_{+0}^{x} dt/\phi(t) < C \int_{+0}^{x} |y'(t)|\, dt = C\,|y(x) - y(+0)|.$$

This, however, contradicts (4).

Since $f(x, b)$ and $f(x, -b)$ are of opposite signs, there exists, corresponding to every $x > 0$, at least one y for which $f(x, y) = 0$. For a fixed $x > 0$, let $y = y^{+}(x)$ be the maximum of such values y. Similarly, let $y = y^{-}(x)$ be the minimum of such values y. Then

(12) $$-b < y^{-}(x) \leqq y^{+}(x) < b, \quad \text{where} \quad 0 < x \leqq a,$$

and

(13) $$y^{-}(x) \to 0, \quad y^{+}(x) \to 0 \quad \text{as} \quad x \to +0.$$

The limit relations (13) are consequences of the fact that $1/f(x, y)$ is bounded for small x outside of any vicinity of $(0, 0)$.

What remains to be shown is that the limit $y(+0)$ must exist. This is obvious if $y(x)$ is monotone for small x, since the existence of $y(x)$ implies, of course, that the curve $y = y(x)$ remains in R. In the remaining case, where $y(x)$ is not monotone for sufficiently small x, the derivative $y'(x)$ must change its sign and therefore must be zero for certain arbitrarily small values of x. In other words, there must exist a sequence $x_1, x_2, \cdots$ satisfying $x_n > x_{n+1} \to 0$ and $f(x_n, y(x_n)) = 0$. But then

(14) $$\min y^{-}(x) \leqq \min y(x) \leqq \max y(x) \leqq \max y^{+}(x), \quad \text{where } x_{n+1} \leqq x \leqq x_n.$$

For if (14) did not hold, it would follow that there exists an $x = x^0$, such that $x_{n+1} < x^0 < x_n$ and, for example,

(15) $$y(x^0) > \max y^{+}(x), \quad \text{where} \quad x_{n+1} \leqq x \leqq x_n.$$

This implies that $f(x, y(x))$ is not zero at $x = x^0$, and so it does not vanish on an open x-interval containing $x = x^0$. Let x^* denote an end-point of this interval. Then $f(x^*, y(x^*)) = 0$; hence, by the definition of $y^{+}(x)$,

(16) $$y(x^*) \leqq y^{+}(x^*), \quad \text{where} \quad x_{n+1} \leqq x^* \leqq x_n.$$

But $y(x)$ is monotone on the interval having the end-point x^*, since $y'(x)$ does not vanish in the interior. Hence at one end-point, say at $x = x^*$,

$$y(x^*) > y(x^0).$$

But this contradicts (15) and (16).

Since (13) and (14) imply (10), the proof of the last italicized statement is complete.

5. In order to prove (I), there remains to be shown that, if $y = y(x)$ is any solution path of (3) through an interior point (x_0, y_0) of (11), then $y(x)$ can be continued for all positive $x (< x_0)$. However, such a continuation could not be made only if $y(x)$ approached the lower or upper boundary of R, as x approaches some positive number. But this is impossible, since $y'(x) > 0$ if $y(x)$ is near b, and $y'(x) < 0$ if $y(x)$ is near $-b$. This establishes (I).

In order to prove (II), let $x_1, x_2, \cdots$ be a sequence of numbers satisfying $a > x_n > x_{n+1} \to 0$. Since $f(x_1, b) < 0$, there exists, on some small interval to the right of x_1, a solution path $y = y_1(x)$ satisfying $y_1(x_1) = b$. By using arguments as above, it is easy to see that this solution can be continued for all x contained in the interval $x_1 \leqq x \leqq a$. Similarly, let $y = y_2(x)$, where $x_2 \leqq x \leqq a$, be a solution of (3) satisfying $y_2(x_2) = b$. Such a solution path $y = y_2(x)$ may be so selected that $y_2(x) \leqq y_1(x)$ holds at every x at which both $y_1(x)$ and $y_2(x)$ are defined (for if $y_2 = y_1$ at some $x = x^0$, then y_2 may be chosen to be identical with y_1 when $x \geqq x^0$). On proceeding in this manner, one obtains a sequence of solution paths $y = y_n(x)$, where $n = 1, 2, \cdots$, such that each of the functions

$$(17) \qquad y_m(x), y_{m+1}(x), \cdots$$

is defined on the interval $x_m \leqq x \leqq a$ and, for every fixed x in this interval, the sequence (17) is monotone non-increasing in n. Hence, the sequence is convergent on this closed interval. But, since $f(x, y)/\phi(x)$ is continuous on the rectangle $x_m \leqq x \leqq a, -b \leqq y \leqq b$, the functions (17) are equicontinuous on the interval $x_m \leqq x \leqq a$. Hence, the convergence of the sequence is uniform on this interval, which implies that the limit function of the sequence is a solution of (3).

This completes the proof of (II).

In order to prove (II bis), suppose that, for every fixed x contained in the interval $0 < x \leqq a$, the function $f(x, y)$ is a non-increasing function of y and that there exist two distinct solutions, say $y = y^1(x)$ and $y = y^2(x)$, defined for all small positive x. Then, by the last italicized statement, (10) holds for $y = y^1(x)$ and for $y = y^2(x)$. Since these solutions are distinct,

there is some positive $x = x^0 (\leqq a)$ at which $y^1(x) - y^2(x) > 0$. By continuity, this inequality holds for all x sufficiently close to x^0. Thus it follows from (3), by subtraction, that $y^1(x) - y^2(x)$ has a non-positive derivative and is, therefore, non-increasing in a vicinity of $x = x^0$. Accordingly, if x is sufficiently near and less than x^0,

(18) $$y^1(x) - y^2(x) \geqq y^1(x^0) - y^2(x^0) > 0.$$

This argument also shows that $y^1(x) - y^2(x)$ is non-increasing for all positive $x \leqq x^0$, and so (18) is valid for all such x. But (18) implies that

$$0 = y^1(+0) - y^2(+0) \geqq y^1(x^0) - y^2(x^0) > 0.$$

This contradiction proves (II bis).

6. As seen above, (i) and (ii) are simple corollaries of (I), (II) and (II bis). In what follows, other applications of the general theorems will be deduced, by considering the Briot-Bouquet equation (1) and its generalization (2) under very relaxed restrictions on the non-linear terms.

(iii) *Let $\lambda \geqq 1$, and let $g(x, y)$ be a real-valued function which is continuous on the rectangle*

(6) $$0 \leqq x \leqq a, \qquad -b \leqq y \leqq b$$

and satisfies condition

(19) $$g(0, y) = o(|y|) \quad \text{as} \quad y \to 0;$$

finally, let p be any positive number. Then (6) *contains a rectangle* (9), *having the property that all those solutions of the differential equation*

(20) $$x^\lambda y' = py + g(x, y)$$

which issue from points of (9) *satisfy* (10).

In fact, if (20) is identified with (3), then (4) is satisfied by $\phi(x) = x^\lambda$, since $\lambda \geqq 1$. Hence, it is sufficient to ascertain that the conditions required for

$$f(x, y) = py + g(x, y)$$

in (I) are satisfied if the rectangle (11) in (I) is replaced by some rectangle (9). But this is obvious from $p > 0$ and from (19). In fact, instead of (19), only

(19 bis) $$|g(0,y)| < p\,|y| \text{ for small } |y|, \qquad (0 < |y| < b'),$$

is needed.

Similarly, (II) implies the following theorem:

(iv) *If the assumptions of* (iii) *are satisfied, but* (20) *is replaced by*

(21) $$x^\lambda y' = -py + g(x,y), \qquad (-p < 0),$$

then there exists at least one solution $y = y(x)$ *which exists for all small* $x > 0$, *and all such solutions satisfy* (10).

REMARK. In order to assure that the solution established by (iv) is *unique*, it is sufficient *to impose on* $g(x,y)$ *the monotony conditions which render* (II bis) *applicable* to the particular case (21) of (3).

QUEENS COLLEGE, FLUSHING, N. Y.
THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] H. Dulac, *Curvas definidas por una ecuación diferencial de primer orden y de primer grado*, Madrid, 1933.

[2] H. Liebmann, "Geometrische Theorie der Differentialgleichungen," *Encyklopaedie der mathematischen Wissenschaften*, III D 8 (1914).

[3] O. Perron, "Beweis für die Existenz von Integralen einer gewöhnlichen Differentialgleichung in der Umgebung einer Unstetigkeitsstelle," *Mathematische Annalen*, vol. 75 (1914), pp. 256-273.

[4] A. Wintner, "Unrestricted Riccatian solution fields" (to appear elsewhere).

CONVERSE LINEARITY CONDITIONS.*

By R. H. Bing.

1. **Introduction.** A point is a *between-point* of the plane set M if it is between two points of M; it is a *middle-point* of M if it bisects some interval having its end points on M. The function $f(x)$ satisfies the *between-point condition* on the set E of x's if for each x_0 of E, the point of the graph of $f(x)$ having x_0 as an abscissa is a between-point of the graph of $f(x)$; $f(x)$ satisfies the *middle-point condition* on E if for each x_0 of E, the point of the graph of $f(x)$ having x_0 as an abscissa is a middle-point of the graph of $f(x)$.

Examples of functions satisfying the middle-point condition on their ranges are

$$f(x) = mx + k \ (a < x < b) \quad \text{and} \quad f(x) = \sin x \ (-\infty < x < \infty).$$

It has been shown [2, p. 253; also see 1] that a continuous function $f(x)$ $(a \leq x \leq b)$ is linear if it satisfies the middle-point condition for values of x between a and b. Theorem 1 is a modification of this result.

Theorem 1. *The continuous function $f(x)$ $(a \leq x \leq b)$ is linear if it satisfies the between-point condition for values of x between a and b.*

Proof. Letting $F(x)$ denote the function whose graph is the interval having end points at $[a, f(a)]$ and $[b, f(b)]$, we consider

$$u(x) = f(x) - F(x) \qquad (a \leq x \leq b). \tag{1}$$

We note that $u(x)$ satisfies the between-point condition for values of x between a and b and that $u(a) = u(b) = 0$. As $u(x)$ is continuous and has a closed range, it takes on a least upper bound on a closed subset E of its range. If x_0 is the maximum element of E and if $u(x_0) \neq 0$, then $u(x)$ does not satisfy the between-point condition at x_0. Therefore

$$u(x) \leq 0 \qquad (a \leq x \leq b). \tag{2}$$

Considering the greatest lower bound of $u(x)$, we find that

$$u(x) \geq 0 \qquad (a \leq x \leq b). \tag{3}$$

* Received February 11, 1946. Presented to the American Mathematical Society, April 27, 1946.

From (2) and (3) we have that

$$(4) \qquad u(x) = 0 \qquad (a \leq x \leq b)$$

and from (1) and (4) that $f(x)$ is linear.

THEOREM 2. *The continuous function $f(x)$ $(a < x < b)$ is linear if $\int_a^b f(x)\,dx$ exists and for each x_0 between a and b there are an $h = h(x_0)$ and a $k = k(x_0)$ such that*

$$(5) \qquad k^2[hf(x_0) - \int_{x_0-h}^{x_0} f(x)\,dx] = h^2[\int_{x_0}^{x_0+k} f(x)\,dx - kf(x_0)]$$
$$(a \leq x_0 - h < x_0 < x_0 + k \leq b).$$

Proof. First, we shall show that there is a continuous function $f'(x)$ $(a \leq x \leq b)$ such that $f'(x) \equiv f(x)$ $(a < x < b)$.

Assume that (b, y_1) and (b, y_2) are limit points of the graph of $f(x)$. Suppose that c and d are values such that $a < c < d < b$. Since $\int_a^b f(x)\,dx$ exists, $\int_x^d f(y)\,dy$ $(a \leq x \leq d)$ is continuous and takes on a maximum M and a minimum m. Let $L(x)$ and $R(x)$ $(a \leq x \leq b)$ be continuous linear functions such that $L(b) = R(b) = (y_1 + y_2)/2$, $L(x) > f(x) > R(x)$ $(c < x < d)$, $L(x) > 0 > R(x)$ $(a < x < c)$, $\int_c^d L(x)\,dx > M$ and $m > \int_c^d R(x)\,dx$. Since $L(x) > f(x)$ $(c < x < d)$ and $\int_x^d L(y)\,dy > M$ $(a \leq x \leq c)$, it follows that

$$(6) \qquad \int_x^d [f(y) - L(y)]\,dy < 0 \qquad (a \leq x < d).$$

Let r and s be values between d and b such that $r < s$, $f(r) > L(r)$ and $f(s) < R(s)$. Denote by x' the largest value for which $f(x) - L(x)$ $(d \leq x \leq s)$ assumes a maximum. Assume that there are values h and k such that (5) is satisfied for x' equal to x_0. We note that if $f(x)$ is linear, then (5) holds for all h and k such that $a \leq x_0 - h < x_0 < x_0 + k \leq b$. Hence,

$$(7) \qquad k^2\{h[f(x') - L(x')] - \int_{x'-h}^{x'} [f(x) - L(x)]\,dx\}$$
$$= h^2\{\int_{x'}^{x'+k} [f(x) - L(x)]\,dx - k[f(x') - L(x')]\}.$$

From (6) and the fact that $f(x') - L(x')$ is a positive maximum for $f(x) - L(x)$ $(d \leqq x \leqq s)$, it follows that the first member of (7) is not negative. Therefore,

$$(8) \qquad \int_{x'}^{x'+k} [f(x) - L(x)]dx \geq k[f(x') - L(x')].$$

Since x' is the largest value for which $f(x) - L(x)$ $(d \leq x \leq s)$ assumes a maximum, it follows from (8) that $x' + k > s$ and $\int_s^{x'+k} f(x)dx > 0$.

Let E be the set of all values x such that $s \leq x \leq b$ and $\int_s^x [f(y) - L(y)]dy \geq 0$. By the methods of the last paragraph we find that b is either a point or a limit point of E. Since $\int_a^b f(x)dx$ exists, it follows that $\int_s^b [f(x) - L(x)]dx \geq 0$. A similar argument leads to the contradiction that $\int_s^b [f(x) - R(x)]dx \leq 0$.

Assume that $f(x) \to \infty$ as $x \to b$. There is a continuous linear function $G(x)$ such that $f(x) > G(x)$ $(c < x < d)$, $G(x) < 0$ $(a < x < d)$, $m > \int_c^d G(x)dx$ and there is a value t between d and b for which $f(t) < G(t)$. Let x' be the largest value for which $f(x) - G(x)$ $(d \leq x < b)$ assumes its minimum. But there are not an h and a k such that (5) is satisfied for x' equal to x_0. Likewise, the assumption that $f(x) \to -\infty$ as $x \to b$ leads to a similar contradiction. Hence, $f(x)$ may be defined so as to be continuous at b. Also, it may be defined so as to be continuous at a. Therefore, there is a continuous function $f'(x)$ $(a \leq x \leq b)$ such that $f'(x) \equiv f(x)$ $(a < x < b)$.

Let $F(x)$ $(a \leq x \leq b)$ be a continuous linear function such that $f'(a) = F(a)$ and $f'(b) = F(b)$. Assume that $f'(x) - F(x)$ is positive for some value between a and b. Let x' be the largest value for which $f'(x) - F(x)$ takes on its maximum. But there are not an h and a k such that (5) is satisfied for x' equal to x_0. Also, the assumption that $f'(x) - F(x)$ is negative for some value between a and b leads to a contradiction.

COROLLARY. *The continuous function $f(x)$ $(a < x < b)$ is linear if $\int_a^b f(x)dx$ exists and if for each x_0 between a and b there is an $h = h(x_0)$ such that*

$$f(x_0) = (1/2h) \int_{x_0-h}^{x_0+h} f(x)dx \qquad (a \leq x_0 - h < x_0 + h \leq b).$$

2. Middle-points of graphs. We shall use Theorem 3 in the next section. Theorem 4 is included because it is of interest for itself.

THEOREM 3. *If (x_0, y_1) and (x_0, y_2) are middle-points of the graph of the continuous function $f(x)$ $(a < x < b)$ and if $y_1 < y_0 < y_2$, then (x_0, y_0) is a middle-point of the graph of $f(x)$.*

Proof. Let $u_1(x)$, $u_0(x)$ and $u_2(x)$ be functions whose graphs are the images of the graph of $f(x)$ under a rotation of an angle of π radians about (x_0, y_1), (x_0, y_0) and (x_0, y_2) respectively. As $y_1 < y_0 < y_2$, we have

$$u_1(x) < u_0(x) < u_2(x) \qquad (2x_0 - b < x < 2x_0 - a). \tag{9}$$

Since (x_0, y_1) and (x_0, y_2) are middle-points of the graph of $f(x)$, there are values x_1, x_2 such that for $i = 1, 2$ we have

$$u_i(x_i) = f(x_i) \qquad [x_0 < x_i < \min(b, 2x_0 - a)]. \tag{10}$$

From (9) and (10) it follows that

$$u_0(x_1) > f(x_1) \quad \text{and} \quad u_0(x_2) < f(x_2). \tag{11}$$

From (11) and the continuity of $f(x)$ and $u_0(x)$, it follows that there is a number x_3 between x_1 and x_2 such that $u_0(x_3) = f(x_3)$. Then (x_0, y_0) bisects the interval from $[x_3, f(x_3)]$ to $[2x_0 - x_3, f(2x_0 - x_3)]$ and is therefore a middle-point of the graph of $f(x)$.

THEOREM 4. *If the continuous function $f(x)$ $(a < x < b)$ satisfies the middle-point condition on its range and for no value c between a and b is either $f(x)$ $(a < x < c)$ or $f(x)$ $(c < x < b)$ linear, then the set of all middle-points of the graph of $f(x)$ consists of the sum of two connected domains plus a subset of their boundaries.*

Proof. The existence of a continuous function $f(x)$ $(a < x < b)$ satisfying the middle-point condition and being nonlinear on every connected subset of its range will be shown in the next section.

First, it will be shown that if (x_0, y_1) and (x_0, y_2) are middle-points of the graph of $f(x)$ and if $y_1 < y_0 < y_2$, then there is a positive number ϵ such that if $|\delta| < \epsilon$, $(x_0 + \delta, y_0)$ is a middle-point of the graph of $f(x)$. Employing the notation and the methods used in the proof of Theorem 3, we have that there are an x_1 and an x_2 such that

$$u_0(x_1) > f(x_1) \qquad [x_0 < x_1 < \min(b, 2x_0 - a)]$$

and

$$u_0(x_2) < f(x_2) \qquad [x_0 < x_2 < \min(b, 2x_0 - a)].$$

From the continuity of $u_0(x)$ and $f(x)$, we find that there is a positive number ϵ such that if $|2\delta| < \epsilon$, then

$$u_0(x_1) > f(x_1 + 2\delta) \qquad [x_0 < x_1 + 2\delta < \min(b, 2x_0 - a)]$$

and

$$u_0(x_2) < f(x_2 + 2\delta) \qquad [x_0 < x_2 + 2\delta < \min(b, 2x_0 - a)].$$

From the continuity of $u_0(x)$ and $f(x)$, there is an x_3 between x_1 and x_2 such that $u_0(x_3) = f(x_3 + 2\delta)$. Then $(x_0 + \delta, y_0)$ is a midpoint of the interval from $[2x_0 - x_3, f(2x_0 - x_3)]$ to $[x_3 + 2\delta, f(x_3 + 2\delta)]$.

There are positive numbers ϵ_1 and ϵ_2 such that if $|\delta| < \epsilon_i$ $(i = 1, 2)$, then $[x_0 + \delta, (y_0 + y_i)/2]$ is a middle-point of the graph of $f(x)$. Applying Theorem 3, we find that a connected domain contains (x_0, y_0) which contains only middle-points of the graph of $f(x)$.

We shall show that if x_0 is a number between a and $(a + b)/2$, then there is more than one middle-point of the graph of $f(x)$ with an abscissa equal to x_0. Assume that (x_0, y) is a middle-point of the graph of $f(x)$ only if $y = f(x_0)$. If h is a positive number such that $a < x_0 - h$, then the midpoint of the segment from $[x_0 - h, f(x_0 - h)]$ to $[x_0 + h, f(x_0 + h)]$ is a middle-point of the graph of $f(x)$ having an abscissa equal to x_0. The ordinate of this midpoint is $f(x_0)$ because (x_0, y) is a middle-point of the graph of $f(x)$ only if $y = f(x_0)$. Hence, $f(x)$ $(a < x < 2x_0 - a)$ is symmetric with respect to $[x_0, f(x_0)]$. As $f(x)$ is continuous at $2x_0 - a$, there is a continuous function $g(x)$ $(a \leq x \leq 2x_0 - a)$ such that $g(x) = f(x)$ for values of x between a and $2x_0 - a$. Since $g(x)$ is not linear, we have by Theorem 1 that there is a number x_1 between a and $2x_0 - a$ such that $[x_1, g(x_1)]$ is not a middle-point of the graph of $g(x)$. Since $g(x)$ is symmetric with respect to $[x_0, g(x_0)]$, we may assume that $x_1 < x_0$. Then $[x_1, f(x_1)]$ is not a middle-point of the graph of $f(x)$. Hence, the assumption that there is only one middle-point of the graph of $f(x)$ with an abscissa equal to x_0 leads to the contradiction that $f(x)$ does not satisfy the middle-point condition at x_1.

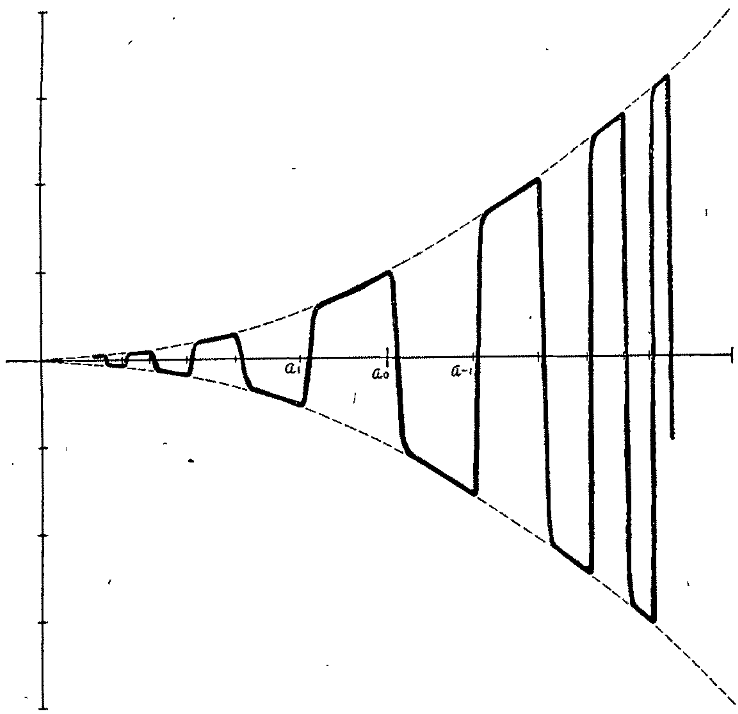
If x_0 is a value between a and $(a + b)/2$, then the nondegenerate connected set of all middle-points of the graph of $f(x)$ with abscissas equal to x_0 is covered by two points plus a domain consisting of middle-points of the graph of $f(x)$. Then the set of middle-points of the graph of $f(x)$ with abscissas less than $(a + b)/2$ consists of a simply connected domain D plus a subset of its boundary. Also, the set of middle-points of the graph of $f(x)$ with abscissas greater than $(a + b)/2$ consists of a simply connected domain D' plus a subset of its boundary. The set of middle-points of the graph of $f(x)$ with abscissas equal to $(a + b)/2$ is a subset of both the boundary of D and the boundary of D'.

3. Functions satisfying the middle-point condition. An example [2, pp. 253-255] has been given of a continuous nonlinear function $f(x)$ $(a < x < b)$ which satisfies the middle-point condition on its range. The graph of this function is the sum of a countable number of straight line intervals. One might wonder whether a continuous function $f(x)$ $(a < x < b)$ which satisfies the middle-point condition on its range could be nonlinear on each subinterval of its range.

THEOREM 5. *There exists a bounded function $F(x)$ $(a \leq x < b)$ having a derivative everywhere on its range, satisfying the middle-point condition for values of x between a and b and being nonlinear on each subinterval of its range.*

Proof. Designate by $a_0, a_1, a_2, \cdots, a_n, \cdots$ the values 1, 3/4, 9/16, $\cdots$, $(3/4)^n, \cdots$ and by $a_{-1}, a_{-2}, \cdots, a_{-n}, \cdots$ the values 5/4, 23/16, $\cdots$, $2-(3/4)^n, \cdots$. Let

$$F(x) = (-1)^n x^2 \qquad [(4a_{n+1} + a_n)/5 \leq x \leq a_n]. \tag{12}$$



For other x such that $0 \leq x \leq 2$, $F(x)$ is defined so as to be nonlinear on each subinterval, so as to have a derivative at every point of its range and so that $|F(x)| \leq x^2$. See the figure.

We shall show that for each x_0 between 0 and 2, there are middle-points (x_0, y_1) and (x_0, y_2) of the graph of $F(x)$ such that $y_1 > x_0{}^2$ and $y_2 < -x_0{}^2$. An application of Theorem 3 will give Theorem 5.

Let W_n be the set of points on the x-axis whose abscissas x satisfy $(4a_{n+1} + a_n)/5 \leq x \leq a_n$. We shall show that if x_0 is a number between 0

and 2, then $(x_0, 0)$ is both a middle-point of $\sum W_{2n}$ and a middle-point of $\sum W_{2n+1}$.

Suppose that $(x_0, 0)$ is not a middle-point of any W_{2n+1}. There is a number j such that $(x_0, 0)$ is between a point of W_{2j+1} and a point of W_{2j-1}. A computation shows that if $j > 0$, then $(x_0, 0)$ bisects an interval having its end points on $W_{2j-1} + W_{2j-2}$ and $\sum_j^\infty W_{2n+1}$ respectively; if $j < 0$, then $(x_0, 0)$ bisects an interval having its end points on $W_{2j+3} + W_{2j+1}$ and $\sum_{-\infty}^j W_{2n-1}$ respectively; if $j = 0$ and $x_0 \leq 1$, then $(x_0, 0)$ bisects an interval from W_{-1} to $W_1 + W_3$; if $j = 0$ and $1 + (7/160)(3/4)^{2m+1} \leq x_0 \leq 1 + (7/32)(3/4)^{2m+1}$ for $m = 0, 1, \cdots$, then $(x_0, 0)$ bisects an interval from W_{2m+1} to W_{-2m-3}. Hence, each point of the x-axis between $(0, 0)$ and $(2, 0)$ is a middle-point of $\sum W_{2n+1}$. Likewise, each such point is a middle-point of $\sum W_{2n}$.

There are points $(x_1, 0)$ and $(x_2, 0)$ of $\sum W_{2n+1}$ such that $x_0 = (x_1 + x_2)/2$. We have by (12) that $[x_0, -(x_1^2 + x_2^2)/2]$ is a middle-point of the graph of $F(x)$. But $y_2 = -(x_1^2 + x_2^2)/2 < -x_0^2$. Likewise, there is a point (x_0, y_1) which is a middle-point of the graph of $f(x)$ and such that $y_1 > x_0^2$. We have by Theorem 3 that $F(x)$ satisfies the middle-point condition for values between 0 and 2.

THEOREM 6. *If each between-point of the graph of $f(x)$ $(a < x < b)$ is a middle-point of this graph and if $f(x)$ $(a \leq x < b)$ has a derivative at $x = a$, then $f(x)$ is linear.*

Proof. There is a continuous linear function $L(x)$ $(a \leq x < b)$ whose graph is tangent to the graph of $f(x)$ at $[a, f(a)]$. Assume that there is an x_0 between a and b such that $f(x_0) \neq L(x_0)$. Let $M(x)$ be the function whose graph is the interval joining $[a, f(a)]$ and $[x_0, f(x_0)]$. As the graph of $L(x)$ is tangent to the graph of $f(x)$ at $[a, f(a)]$, there is an x_1 such that $|L(x) - f(x)| < |M(x) - f(x)|$ for values of x between a and x_1. Then $[(2a + x_1)/3, M([2a + x_1]/3)]$ is a limit point of between-points of the graph of $f(x)$ $(a < x < b)$ but it is not a limit point of middle-points of this graph.

THEOREM 7. *Suppose that AB is an interval from the point A to the point B. There exist two totally disconnected closed subsets H and K of AB such that $H \cdot K = A + B$ and such that each point of $AB - (A + B)$ is both a middle-point of $H - (A + B)$ and a middle-point of $K - (A + B)$.*

Proof. First, we shall describe a totally disconnected closed subset $S(I)$ of an interval I. Let E_1 be the segment consisting of the middle one-fifth of I. Then $I - E_1$ is the sum of two intervals. Let E_2 be the sum of two segments

each of which is the middle one-tenth of one of these intervals. In general, let E_{i+1} be the sum of all segments s such that s is the middle $1/(5\cdot 2^i)$ of a component of $I-(E_1+E_2+\cdots+E_i)$. The desired point set $S(I)$ is $I-(E_1+E_2+\cdots)$.

We shall note some properties of $S(I)$ that make it useful in proving Theorem 7. It is closed and totally disconnected. We shall show that if PQ is a subinterval of I such that P is an end point of I, then

(13) $$\text{measure } S(I)\cdot PQ \geq 1/2 \text{ length } PQ.$$

Since $E_1+E_2+\cdots$ is dense in I, (13) is true if it holds for each point Q of $E_1+E_2+\cdots$. Assume that (13) is false. Let $\bar{E}_j$ be the first element of $\bar{E}_1, \bar{E}_2, \cdots$ containing a point Q such that (13) does not hold. For convenience, we shall assume that the length of I is 1. Then the measure of $E_1+E_2+\cdots$ is $1/5+1/(5\cdot 2)+\cdots=2/5$ and the measure of $S(I)$ is $1-2/5=3/5$. The length of each component of E_i is $1/(5\cdot 4^{i-1})$ and the length of each component of $I-(E_1+E_2+\cdots+E_i)$ is

$$(1/2^i)\{1-[1/5+1/(5\cdot 2)+\cdots+1/(5\cdot 2^{i-1})]\}$$
$$=(3\cdot 2^i+2)/(5\cdot 4^i).$$

The measure of the common part of $S(I)$ and a component of $I-(E_1+E_2+\cdots+E_i)$ is $(1/2^i)(3/5)$. If Q is a point of $\bar{E}_1$, the measure of $S(I)\cdot PQ$ is $3/10$ and the length of PQ is no more than $3/5$. Therefore, $\bar{E}_1$ is not $\bar{E}_j$. If Q' is the last point of $P+\bar{E}_1+\bar{E}_2+\cdots+\bar{E}_{j-1}$ on PQ in the order from P to Q, then

$$\text{measure } S(I)\cdot PQ' \geq 1/2 \text{ length } PQ'.$$

But,

$$\text{measure } S(I)\cdot QQ' = (1/2^i)(3/5)$$

and

$$\text{length } QQ' \leq (3\cdot 2^i+2)/(5\cdot 4^i)+1/(5\cdot 4^{i-1})$$
$$=(3\cdot 2^i+6)/(5\cdot 4^i).$$

Then,

$$\text{measure } S(I)\cdot QQ' \geq 2^i/(2^i+2) \text{ length } QQ'.$$

Hence, (13) holds for all Q in I. Using a similar line of thought, we find that if UV is a subarc of I not intersecting $E_1+E_2+\cdots+E_i$ and U is a point of $\bar{E}_i$, then

(14) $$\text{measure } S(I)\cdot UV \geq 2/3 \text{ length } UV.$$

Each middle-point of I is a middle-point of $S(I)$. To see this, let I' and $S(I')$ be the images of I and $S(I)$ under a rotation of π radians about a middle-point R of I. If PQ is the common part of I and I', we have by (13)

that both the measure of $S(I) \cdot PQ$ and the measure of $S(I') \cdot PQ$ is as much as 1/2 length PQ. Hence $S(I)$ intersects $S(I')$ and R is a middle-point of $S(I)$.

If K is an interval, let $S(K)$ denote the image of $S(I)$ under a similarity transformation of I into K. It will be shown that if I and K are two intervals that intersect in an interval L and neither I nor K is as much as five times as long as the other, then $S(I)$ intersects $S(K)$ on a set of positive measure. It will follow that if M and N are subintervals of a straight line and neither M nor N is as much as five times as long as the other, then each middle-point of $M + N$ is a middle-point of $S(M) + S(N)$.

If neither I nor K is a subset of the other, then one end point of L is an end point of I and the other end point of L is an end point of K. By (13) we have that

$$\text{measure } S(I) \cdot L \geq 1/2 \text{ length } L$$

and that

$$\text{measure } S(K) \cdot L \geq 1/2 \text{ length } L.$$

As $S(I) \cdot L$ and $S(K) \cdot L$ are closed, they intersect on a set of positive measure.

If K is a subset of I, let E_i be the first element of $E_1, E_2, \cdots$ intersecting K. Since

$$\text{length } E_i \leqq 1/5 \text{ length } I < \text{length } K,$$

K is not a subset of E_i. Let L' be a maximal subinterval of L containing no point of E_i. By (14) we have that

$$\text{(15)} \qquad \text{measure } S(I) \cdot L' \geq 2/3 \text{ length } L'.$$

But one end point of L' is an end point of L and therefore of K. By (13) we have that

$$\text{(16)} \qquad \text{measure } S(K) \cdot L' \geq 1/2 \text{ length } L'.$$

Now (15) and (16) give that $S(I)$ and $S(K)$ intersect on a set of positive measure.

Suppose that AB is the interval from $(0,0)$ to $(2,0)$. As in Theorem 5, designate by $a_0, a_1, \cdots, a_n, \cdots$ the values $1, 3/4, \cdots, (3/4)^n, \cdots$ and by $a_{-1}, a_{-2}, \cdots, a_{-n}, \cdots$ the values $5/4, 23/16, \cdots, 2 - (3/4)^n, \cdots$. Let M_n be the set of points $(x, 0)$ such that $(7a_{n+1} + a_n)/8 \leq x \leq a_n$, let $H - (A + B) = \sum S(M_{2n})$ and let $K - (A + B) = \sum S(M_{2n+1})$.

If $(x_0, 0)$ is between two points of M_{2n+1}, it is a middle-point of $S(M_{2n+1})$. If $(x_0, 0)$ is a point of $AB - (A + B)$ but is not between any two points of any M_{2n+1}, there is a j such that $(x_0, 0)$ is between a point of M_{2j+1} and a

point of M_{2j-1}. A computation shows that if $j < 0$, then $(x_0, 0)$ is a middle-point of $M_{2j+1} + M_{2j-1} + M_{2j-3}$ and is therefore a middle-point of $S(M_{2j+1}) + S(M_{2j-1}) + S(M_{2j-3})$; if $j > 0$, then $(x_0, 0)$ is a middle-point of $S(M_{2j+3}) + S(M_{2j+1}) + S(M_{2j-1})$; if $j = 0$ and $x_0 \leq 1$, then $(x_0, 0)$ is a middle-point of $S(M_3) + S(M_1) + S(M_{-1})$; if $j = 0$ and $1 + (7/256)(3/4)^{2m+1} \leq x \leq 1 + (7/32)(3/4)^{2m+1}$ for $m = 0, 1, \cdots$, then $(x_0, 0)$ is a middle-point of $S(M_{2m+1}) + S(M_{-2m-3})$. Hence, each point of $AB - (A + B)$ is a middle-point of $K - (A + B)$. Likewise, it can be shown that each point of $AB - (A + B)$ is a middle-point of $H - (A + B)$.

THEOREM 8. *There exists a bounded function $F(x)$ $(a < x < b)$ satisfying the middle-point condition and having a derivative on its range such that $F(x)$ is nonlinear on each subinterval of its range and every between-point of the graph of $F(x)$ is a middle-point of this graph.*

Proof. The function

$$f(x) = \frac{1}{x - x^2} \sin \frac{1}{x - x^2} \qquad (0 < x < 1)$$

satisfies all of the conditions of this theorem except the one of boundedness.

Let AB be the interval from $(0, 0)$ to $(1, 0)$ and let H and K be totally disconnected closed subsets of AB such that $H \cdot K = A + B$ and each point of $AB - (A + B)$ is both a middle-point of $H - (A + B)$ and a middle-point of $K - (A + B)$. We define $F(x)$ $(0 < x < 1)$ to be a function having a derivative on its range and such that $|F(x)| \leq x$, the projection on the x-axis of the points of the graph of $F(x)$ for which $F(x) = x$ is the set $H - (A + B)$, the projection on the x-axis of the points of the graph of $F(x)$ for which $F(x) = -x$ is the set $K - (A - B)$ and $F(x)$ is nonlinear on each subinterval. Since each point between $(0, 0)$ and $(1, 1)$ is a middle-point of the graph of $F(x)$ and each point between $(0, 0)$ and $(1, -1)$ is a middle-point of the graph of $F(x)$, it follows by Theorem 3 that every between-point of the graph of $F(x)$ is a middle-point of this graph.

THE UNIVERSITY OF TEXAS.

REFERENCES.

1. E. F. Beckenbach, "On a characteristic property of linear functions," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 923-930.

2. R. Courant and D. Hilbert, *Methoden der mathematischen Physik*, vol. II, Berlin, 1937.

A DENSITY THEOREM FOR POWER SERIES.*

By R. P. Boas, Jr.

Let $f(z) = \sum_{n=0}^{\infty} c_n z^n$ have $|z| = 1$ as its circle of convergence.

Theorem. *Suppose that $\{c_n\}$ is not bounded, but that there are numbers M, L and β, $\beta > 1$, such that $|c_{\lambda_n}| \leqq M$ when $|\lambda_n - n\beta| < L$. Then $f(z)$ cannot be majorized by an integrable function $\psi(\theta)$ in any sector of the unit circle of angle exceeding $2\alpha = 2\pi(1 - 1/\beta)$; that is, we cannot have*

$$|f(re^{i\theta})| \leqq \psi(\theta), \qquad 0 \leqq r < 1, \tag{1}$$

in any θ-interval of length exceeding 2α. The same conclusion holds if c_n does not approach zero, but $c_{\lambda_n} \to 0$ when $|\lambda_n - n\beta| < L$. In particular, in either case $f(z)$ must have a singular point in every arc of $|z| = 1$ of length exceeding 2α.

If we have $c_{\lambda_n} = 0$ or even $c_{\lambda_n} = O(\tau^{-n})$, $\tau > 1$, the existence of a singular point, but not the impossibility of (1), follows by Pólya's gap theorem[1] merely from $\lim_{n\to\infty} \lambda_n/n = \beta$, with no further hypothesis on $\{c_n\}$. For an illustration, we can take $f(z) = (1 - z^2)^{1/2}$, where $\lambda_n = 2n + 1$, $M = 0$, $\beta = 2$.

The case where $\lambda_{n+1} - \lambda_n \to \infty$ and (1) fails in every sector was given by Duffin and Schaeffer.[2] Our theorem corresponds to theirs as Pólya's gap theorem corresponds to Fabry's. Our proof is an adaptation of that given by Duffin and Schaeffer for their theorem.

We can assume without loss of generality that (1) is satisfied in $\pi - \gamma \leqq \theta \leqq \pi + \gamma$, $\gamma > \alpha$. Then we have

$$2\pi i c_n = \int_C w^{-n-1} f(w)\, dw,$$

where C is the curve made up of the arc of $|z| = 1$ from $\arg z = \pi - \gamma$ to $\arg z = \pi + \gamma$, the segments of $\arg z = \pi \pm \gamma$ from $r = \rho$ to $r = 1$ $(\rho < 1)$, and the arc of $|z| = \rho$ from $\arg z = -(\pi - \gamma)$ to $\arg z = \pi - \gamma$. Thus

* Received November 16, 1945.

[1] See, for example, N. Levinson, *Gap and Density Theorems*, New York, 1940, p. 89.

[2] R. J. Duffin and A. C. Schaeffer, "Power series with bounded coefficients," *American Journal of Mathematics*, vol. 67 (1945), pp. 141-154; 153.

(2) $$2\pi c_n = F_1(n) + F_2(n),$$
where

$$F_1(z) = \rho^{-z}\int_{-(\pi-\gamma)}^{\pi-\gamma} e^{iz\theta} f(\rho e^{i\theta})\,d\theta + ie^{i(\pi-\gamma)z}\int_\rho^1 t^{-z-1} f(te^{i(\pi-\gamma)})\,dt$$
$$- ie^{i(\pi-\gamma)z}\int_\rho^1 t^{-z-1} f(te^{-i(\pi-\gamma)})\,dt,$$
$$F_2(z) = \int_{\pi+\gamma}^{3\pi-\gamma} e^{-iz\theta} f(e^{i\theta})\,d\theta.$$

When $x \to \infty$ through real positive values, $F_2(x) \to 0$. Hence $\{F_1(\lambda_n)\}$ is bounded if $\{c_{\lambda_n}\}$ is bounded, and $F_1(\lambda_n) \to 0$ if $c_{\lambda_n} \to 0$. $F_1(z)$ is an entire function of order 1 and type $\pi - \gamma - \log \rho$, which can be made less than $\pi - \alpha = \pi/\beta$ by choosing ρ near enough to 1. A result of Duffin and Schaeffer[3] states that, if $|\lambda_n - n\beta| < L$ and the type of $F_1(z)$ is less than π/β, then $F_1(x)$ is bounded for real positive x if $\{F_1(\lambda_n)\}$ is bounded, and $F_1(x) \to 0$ if $F_1(\lambda_n) \to 0$. By (2), this means that $\{c_n\}$ is bounded or $c_n \to 0$, respectively, contradicting the hypotheses of the theorem.

BROWN UNIVERSITY.

[3] Duffin and Schaeffer, *op. cit.*, pp. 142-143.

A SOLUTION THEORY OF THE MÖBIUS INVERSION.*

By Aurel Wintner.

1. If the summation index m runs through all divisors of n (including the divisor $m = 1$ and, if $n \neq 1$, the divisor $m = n$), then the linear transformation

$$Y_n = \sum_{m|n} X_m, \qquad (n = 1, 2, \cdots) \tag{1}$$

of the infinite sequence $(X_1, X_2, \cdots)$ into $(Y_1, Y_2, \cdots)$ has a unique inverse. In fact, since the n-th of the equations (1) does not contain $X_{n+1}, X_{n+2}, \cdots$ and contains X_n, the infinite system of equations (1) can be solved recursively. The explicit form of the resulting inversion of (1) is known to be the linear transformation

$$X_n = \sum_{m|n} \mu(n/m) Y_m, \qquad (n = 1, 2, \cdots). \tag{2}$$

Here $\mu(1), \mu(2), \cdots$ denotes the sequence of the absolute constants $\mu(k) = \pm(\frac{1}{2} \pm \frac{1}{2})$, defined by the recursive formula and the initial condition

$$\sum_{d|k} \mu(d) = 0 \quad \text{if} \quad k > 1 \quad \text{and} \quad \mu(1) = 1 \tag{3}$$

respectively. An equivalent definition is that

$$\mu(k) = 0 \quad \text{or} \quad \mu(k) = (-1)^\nu \tag{3 bis}$$

according as k is not square-free or is the product of exactly $\nu = \nu(k)$ distinct primes (with the understanding that $k = 1$ belongs to the second case, with $\nu(1) = 0$).

Whether n is or is not square-free, let $\nu(n)$ denote the number of its distinct prime divisors (e. g., $\nu(12) = 2$). Then it is easily realized that $2^{\nu(n)}$ is the number of the square-free divisors of n. Hence, if $\tau(n)$ denotes the number of all divisors of n, then

$$1 \leqq 2^{\nu(n)} \leqq \tau(n) \tag{4}$$

and, since $\nu(2^k) = \nu(2)$ but $\tau(2^k) = k + 1$,

$$\limsup_{n \to \infty} \tau(n)/2^{\nu(n)} = \infty. \tag{5}$$

* Received January 23, 1946.

In view of the sieve process, the infinite matrix defining the linear substitution (1) may be described as follows: The n-th *column* consists of the periodic sequence

$$(6) \qquad (0)_{n-1}, 1, (0)_{n-1}, \cdots, \quad (0)_{n-1}, 1, \cdots,$$

of period n, where $(0)_{n-1}$ denotes the block $0, \cdots, 0$ of $n-1$ consecutive zeros (which is missing when $n=1$; that is, every element of the first column is 1). Correspondingly, it is seen from (2) that the unique inverse of this matrix results by writing the elements of the sequence

$$(7) \qquad (0)_{n-1}, \mu(1), \ (0)_{n-1}, \mu(2), \ (0)_{n-1}, \mu(3), \cdots$$

into the n-th *column* (so that the sequence $\mu(1), \mu(2), \cdots$ forms the first, the sequence $0, \mu(1), 0, \mu(2), \cdots$ the second, $\cdots$ column of the inverse matrix). Hence, if **E** ("Eratosthenes") and **M** ("Möbius") denote the infinite matrices representing the transposed matrices of the linear substitutions (1) and (2) respectively, then the linear substitutions assigned by **E** and **M** are

$$(8) \qquad \mathbf{E}: \ \sum_{m=1}^{\infty} x_{nm} = y_n, \qquad (n = 1, 2, \cdots)$$

and

$$(9) \qquad \mathbf{M}: \ \sum_{m=1}^{\infty} \mu(m) y_{nm} = x_n, \qquad (n = 1, 2, \cdots).$$

In fact, it is clear that the sequences (6), (7) represent the n-th *rows* of the matrices of (8), (9) respectively.

2. Since (1) and (2) are reciprocal mates, and since (8) and (9) are the transposed systems of (1) and (2) respectively, (8) and (9) are *formal* reciprocal mates. In fact, (9) represents the classical "Möbius inversion" of (8).

However, the reservation just italicized is essential indeed. It is true that reciprocation and transposition are commutable in case of finite matrices, and it is also true that the matrices of the reciprocal substitutions (1), (2), being recursive, can be reduced to finite matrices. But this does not insure that the commutability of reciprocation and transposition remains legitimate, since the transposed matrices, that is, the matrices **E**, **M** defined by (8), (9), cannot be reduced to finite matrices.

Actually, very little seems to be known as to the legitimacy of the Möbius inversion, (9), of (8). In fact, all that I find in the literature is contained in a remark of Hardy and Wright ([3], p. 237), according to which (9) is sure to be implied by (8) if

$$\sum_{n=1}^{\infty} \tau(n) \mid x_n \mid < \infty.$$

For the sake of completeness, it will be verified below that, without much additional labor, this criterion can be refined to

$$\sum_{n=1}^{\infty} 2^{\nu(n)} \mid x_n \mid < \infty,$$

which, in view of (4) and (5), is an actual improvement. But *any* criterion of this type is of a trivial nature by necessity.

In addition, any criterion of this type supplies the answer to a question which seems to be quite artificial. In fact, if (9) is thought of as solving the linear equations (8), then what appears to be natural is to subject the data y_n, rather than the unknowns x_n, to restrictions under which the formal solution (9) of (8) is legitimate. However, it turns out that such restrictions cannot exist, simply because the homogeneous equations

$$\sum_{m=1}^{\infty} x_{nm} = 0, \qquad (n = 1, 2, \cdots) \tag{10}$$

possess solutions $(x_1, x_2, \cdots) \neq (0, 0, \cdots)$. But a result of Haar ([2], p. 178) states that

$$\text{either} \quad \sum_{n=1}^{\infty} \mid x_n \mid = 0 \qquad \text{or} \quad \sum_{n=1}^{\infty} \mid x_n \mid = \infty$$

must hold for any solution $(x_1, x_2, \cdots)$ of the homogeneous equations (10). And all of this indicates that the only natural approach to a solution theory of (8) consists in inquiring, not into arbitrary solutions $x = (x_1, x_2, \cdots)$ of (8), but only into the solutions satisfying

$$\sum_{n=1}^{\infty} \mid x_n \mid < \infty; \tag{11}$$

solutions which will be called *regular*.

The purpose of the present paper is to develop the few facts which happen to be true in a general solution theory from this point of view, which corresponds to the principles initiated by Hilbert (in his "bounded" case, where $\Sigma \mid x_n \mid < \infty$ is replaced by $\Sigma \mid x_n \mid^2 < \infty$; cf. the methodical considerations of Hellinger-Toeplitz [4] and Helly [5]). It turns out, among other things, that the range of the "solution theory" is by no means identical with the range of the "Möbius theory." For instance, while it is true that $\mathrm{E}x = y$ cannot possess more than one *regular* solution $x = x(y)$, it is possible that

the latter exists but is not supplied by Möbius' inversion $x = \mathrm{M}y$, the latter being non-existent (that is, the series (9) representing the components of the vector $\mathrm{M}y$ become divergent, although (11) is satisfied).

3. Since E, M result by transposing the matrices of the recursive equations (1), (2), all elements of E, M below the diagonal vanish. Hence, the infinite series defining the elements of the matrix products EM, ME are finite sums, and so both EM and ME exist. Actually, both EM and ME represent the unit matrix. This is readily verified from (3) and from the fact that the sequences (6), (7) are the n-th rows of E, M respectively. Nevertheless, M cannot be denoted by E^{-1} (or E by M^{-1}), since the situation is as follows:

(I) *The matrix* M *is both a right-hand reciprocal and a left-hand reciprocal of the matrix* E. *Furthermore,* M *is the only left-hand reciprocal of* E. *In addition,* E *is the only left-hand reciprocal of* M. *However,* M *is not the only right-hand reciprocal of* E. *In addition,* E *is not the only right-hand reciprocal of* M.

In order to prove this, suppose first that there exist two matrices, M_1 and M_2, for which $\mathrm{M}_1\mathrm{E}$ and $\mathrm{M}_2\mathrm{E}$ become the unit matrix. Then $(\mathrm{M}_1 - \mathrm{M}_2)\mathrm{E}$ is the zero matrix. Hence, if $\mathrm{M}_1 - \mathrm{M}_2$ is not the zero matrix, and if c denotes one of its rows containing at least one non-vanishing element, then $c = (c_1, c_2, \cdots)$ is a non-trivial solution of the homogeneous equations belonging to the *transposed* matrix of E. But (1) shows that these homogeneous equations are

$$0 = \sum_{m|n} c_m \equiv c_1 + \cdots + c_n, \qquad (n = 1, 2, \cdots),$$

and imply, therefore, that $c_1 = 0$, $c_2 = 0, \cdots$. This contradiction proves that $\mathrm{M}_1 = \mathrm{M}_2$.

The uniqueness of the left-hand reciprocal of M follows by a repetition of this argument. This repetition is possible, since the homogeneous equations belonging to the recursive systems (2) have no non-trivial solution.

Similarly, in order to prove that M is not the only right-hand reciprocal of E, it is sufficient to show that the homogeneous equations, $\mathrm{E}x = 0$, possess a non-trivial solution $x = (x_1, x_2, \cdots)$. But such a solution is, for instance, $x_n = \mu(n)/n$, since (10) is then satisfied in view of the well-known relations

$$\sum_{m|n} \mu(m)/m = 0, \qquad (n = 1, 2, \cdots) \tag{12}$$

(Kluyver). Since (12) involves the Prime Number Theorem, it is worth

mentioning that other non-trivial solutions x of $\mathrm{E}x = 0$ may be constructed "elementarily," namely, by ordinary Fourier analysis (in this regard, cf. Rajchman [6]).

Finally, the last assertion of (I) follows if it is shown that the homogeneous system $\mathrm{M}y = 0$ has a non-trivial solution y. But (9) shows that this homogeneous system is

$$(13) \qquad \sum_{m=1}^{\infty} \mu(m) y_{nm} = 0, \qquad (n = 1, 2, \cdots),$$

which, in view of the case $n = 1$ of (12), is satisfied by $y_1 = 1/1$, $y_2 = 1/2$, $y_3 = 1/3, \cdots$.

4. Since the n-th component of the vector $\mathrm{E}x$, where $x = (x_1, x_2, \cdots)$, is the infinite series

$$\sum_{m=1}^{\infty} x_{nm},$$

$\mathrm{E}x$ exists if and only if this series converges for $n = 1, 2, \cdots$. Correspondingly, the assertion that an x is a solution of $\mathrm{E}x = y$ will always imply that this series converges for every n. Clearly, (11) is sufficient for the existence of $\mathrm{E}x$.

In view of (3 bis), these remarks remain valid if x, $\mathrm{E}x$, $\mathrm{E}x = y$ are replaced by y, $\mathrm{M}y$, $\mathrm{M}y = x$ respectively.

A *general* theory of *unrestricted* Möbius solutions is precluded by the following facts:

(II) *If the vector x is a solution of $\mathrm{E}x = y$ (for a given y), then Möbius' vector $\mathrm{M}y = \mathrm{M}(\mathrm{E}x)$, instead of being the vector $(\mathrm{ME})x$ (which, by (I), is the solution x),*

(i) *need not exist;*

(ii) *may exist without being x.*

The proof of (i) will here be omitted, since an assertion much sharper than (i) is contained in the fact (IV) to be proved below.

As to the remaining assertion of (II), it is sufficient to observe that, if y is the zero vector, then $\mathrm{M}y$ exists and is the zero vector; and that the corresponding system $\mathrm{E}x = y$, that is, $\mathrm{E}x = 0$, possesses non-trivial solutions x.

All that happens in this proof of (ii) is that $\mathrm{M}y$ is *a* solution x of $\mathrm{E}x = y$, though it is not the *given* solution. Another possibility is that $\mathrm{M}y$ is a vector not representing *any* solution:

(III) *If Möbius' $\mathrm{M}y$ exists (for a given y), it need not represent a solution x of $\mathrm{E}x = y$.*

In fact, if $y_n = 1/n$ in $y = (y_1, y_2, \cdots)$, then (13) is satisfied, that is, $\mathrm{M}y$ exists and is the vector 0. Hence, if $\mathrm{M}y$ were a solution x of $\mathrm{E}x = y$, it would follow that $\mathrm{E}0 = y$. But this is contradicted by $y_n = 1/n$.

Incidentally, it remains undecided whether there exists a suitable y for which $\mathrm{E}x = y$ has no solution x. All that is clear is that $\mathrm{E}x = y$ can never have a *unique* solution (simply because a solution of $\mathrm{E}x = y$ plus any solution of $\mathrm{E}x = 0$ is a solution of $\mathrm{E}x = y$).

5. The first of the statements of (II) can be refined as follows:

(IV) *There exist vectors y for which the system $\mathrm{E}x = y$ has a regular solution x, although Möbius' solution $\mathrm{M}y$, instead of representing this x, does not exist.*

In other words, not even the restriction (11) can prevent the case (i) of (II). This will be proved in **6.**

As mentioned after (III), it remains undecided whether or not every vector y is representable, in terms of a suitable x, in the form $y = \mathrm{E}x$. On the other hand, it is easy to see that the answer is in the negative if x is restricted by (11). In fact, if $y = (y_1, y_2, \cdots)$ is so chosen as to violate

$$y_n \to 0 \quad \text{as} \quad n \to \infty, \tag{14}$$

then $\mathrm{E}x = y$ cannot have a regular solution $x = x(y)$. In order to see that (14) is a necessary condition for the existence of a regular solution x, it is sufficient to observe that the representation (8) of $\mathrm{E}x = y$ implies the estimate

$$| y_n | \leqq \sum_{m=1}^{\infty} | x_{nm} | \leqq \sum_{k=n}^{\infty} | x_k |,$$

from which the relation (14) follows if (11) is assumed.

In the same direction lies the following fact:

(V) *For a given y, the system $\mathrm{E}x = y$ has either no regular solution or a unique regular solution $x = x(y)$.*

If the assertion of (V) were false (for some fixed y), it would follow by subtraction that the homogeneous system $\mathrm{E}x = 0$ has a regular solution x distinct from $x = 0$. But this contradicts Haar's result, quoted after (10).

6. The proof of (IV) depends on the arithmetical function

$$\phi_n(m) = \sum_{d|n}^{d \leqq m} \mu(d) \tag{15}$$

of the two independent variables n, m (the summation index, d, runs over those divisors of n which do not exceed m).

It is clear from (15) that

$$\phi_n(n) = \phi_n(n+1) = \phi_n(n+2) = \cdots = \begin{cases} 0 \text{ if } n > 1, \\ 1 \text{ if } n = 1, \end{cases} \tag{16}$$

by (3). It is also clear from (15) that

$$\phi_{n+k}(m) = \phi_n(m) \quad \text{if} \quad k = 1 \cdot 2 \cdots m = m!. \tag{17}$$

The periodicity condition (17) implies that $\phi_n(m)$ is a bounded function of n, *if m is fixed*. On the other hand, (16) implies that $\phi_n(m)$ is a bounded function of m, *if n is fixed*. If an assertion of Bruns ([1], p. 132) concerning Möbius' inversion were correct, it would follow that the function $\phi_n(m)$ is bounded *uniformly* in n and m together. However, it was recently shown ([8], § 9–§ 11) that this is not true, since

$$\limsup_{n\to\infty,\ m\to\infty} |\phi_n(m)| = \infty. \tag{18}$$

Trivial upper estimates are

$$|\phi_n(m)| \leqq m \tag{19}$$

and

$$|\phi_n(m)| \leqq 2^{\nu(n)}. \tag{20}$$

In fact, since $\nu(n)$ denotes the number of the distinct prime divisors of n, the number of all square-free divisors of n is $2^{\nu(n)}$, and so it is clear from (3 bis) that

$$\sum_{d|n} |\mu(d)| = 2^{\nu(n)}. \tag{21}$$

But (21) and (15) imply (20). On the other hand, (19) is clear from (3 bis) and (15).

In order to deduce (IV) from (18), let

$$\sum_{n=1}^{\infty} x_n \tag{22}$$

be an absolutely convergent series. Then the same is true of all of the series (8). In particular, $y = \mathrm{E}x$ defines a vector y. For this y, the vector x is a solution of $\mathrm{E}x = y$. On the other hand, substitution of (8) into (9) shows that the first component of the vector $\mathrm{M}y$ is the repeated series

$$\sum_{n=1}^{\infty} \mu(n) \sum_{k=1}^{\infty} x_{nk}. \tag{23}$$

Hence, in order to prove (IV), it is sufficient to show that (11) does not imply the convergence of this repeated series.

Let s_m denote the m-th partial sum of the exterior summation in (23), that is, let

$$s_m = \sum_{n=1}^{m} \sum_{k=1}^{\infty} \mu(n) x_{nk}. \tag{24}$$

This can be rearranged into

$$s_m = \sum_{j=1}^{\infty} x_j \sum_{nk=j}^{n \leqq m} \mu(n), \tag{25}$$

where the interior summation is extended over those positive integers not exceeding m corresponding to which there exists a positive integer k satisfying $nk = j$. This means that n runs through those divisors of j which do not exceed m. Accordingly,

$$s_m = \sum_{j=1}^{\infty} x_j \sum_{d \mid j}^{d \leqq m} \mu(d). \tag{26}$$

Hence, if j is replaced by n, it is seen from (15) and (16) that

$$s_m = \sum_{n=1}^{\infty} \phi_n(m) x_n, \qquad (m = 1, 2, \cdots). \tag{27}$$

Since the absolute constants (15) defining the matrix of the linear substitution (27) satisfy (18), an application of the general norm-principle (Lebesgue-Toeplitz; cf., e. g., [7]) shows that there exist absolutely convergent series (22) for which the sequence $s_1, s_2, \cdots$ becomes divergent. This proves (IV), since s_m is the m-th partial sum of (23).

7. According to (V), the system $\mathrm{E}x = y$ has either no or just one regular solution $x = x(y)$. There arises the need for *practicable* criteria which, when applied to the *data* $y_1, y_2, \cdots$ of the system $\mathrm{E}x = y$, distinguish between the two cases. In this regard, some information is contained in the following theorem:

(VI) *In order that the data $y_1, y_2, \cdots$ be such that the system $\mathrm{E}x = y$, where $y = (y_1, y_2, \cdots)$, has a regular function x (which, by (V), is then unique),*

(i) *the condition $\lim_{n \to \infty} y_n = 0$ is necessary,*

(ii) *the convergence of the series $\sum_{n=1}^{\infty} y_n$ is not necessary,*

(iii) *the condition* $\sum_{n=1}^{\infty} | y_n | < \infty$ *is not sufficient,*

(iv) *the existence of an* $\epsilon > 0$ *satisfying* $\sum_{n=1}^{\infty} n^\epsilon | y_n | < \infty$ *is sufficient.*

Needless to say, it is implied by (iii) that

(i*)–(ii*) *the conditions of* (i) *and* (ii) *are not sufficient,*

and by (ii) that

(iii*)–(iv*) *the conditions of* (iii) *and* (iv) *are not necessary.*

Of the four assertions of (VI), only (i) and (iv) are "criteria" (of the "practicable" type). Actually, (i) is a triviality, verified after (14), and (iv) can be improved as follows:

(iv bis) *If a given* $y = (y_1, y_2, \cdots)$ *satisfies the condition*

$$\sum_{n=1}^{\infty} 2^{\nu(n)} | y_n | < \infty,$$

then $\mathrm{E}x = y$ *has a regular solution* x.

Here $2^{\nu(n)}$ denotes the arithmetical function occurring in that improvement of the Hardy-Wright criterion which was mentioned at the beginning of **2**. However, (iv bis) is quite different from the Hardy-Wright criterion and its improvement, since (iv bis) is a criterion of the "practicable" type, imposing the $2^{\nu(n)}$-condition on the *data* y_n, rather than on the *unknowns* x_n, of the system $\mathrm{E}x = y$. Cf. (VIII) and (X bis) below.

As is well-known, the number, $\nu(n)$, of the distinct prime divisors of n satisfies the estimate

(28) $$\nu(n) = O(\log n)/\log \log n.$$

Hence,

(29) $$2^{\nu(n)} = O(n^\epsilon)$$

holds for every $\epsilon > 0$. Consequently, in order to prove (iv), it is sufficient to verify (iv bis). This will not be done here, since a refinement of (iv bis) is contained in (VIII) below.

In order to prove (ii), let $x_1, x_2, \cdots$ be a sequence of positive numbers satisfying (11) and

(30) $$\sum_{n=1}^{\infty} \tau(n) \, | x_n | = \infty$$

(such sequences exist, since $\tau(n)$, the number of all divisors of n, does not

remain bounded as $n \to \infty$). Since (11) is assumed, it is possible to define by (8) the components of a vector $y = (y_1, y_2, \cdots)$. For this y, the system $Ex = y$ admits the solution $x = (x_1, x_2, \cdots)$, and this x is a regular solution, since it satisfies (11). However, since every x_n was chosen to be positive, it is clear from (8) and from the definition of $\tau(n)$ that

$$\sum_{n=1}^{\infty} y_n = \sum_{n=1}^{\infty} \tau(n) x_n.$$

Hence it is seen from (30), where $x_n > 0$, that the proof of (ii) is complete.

8. Of the four assertions of (VI), only (iii) remains to be considered. It is worth while to formulate (iii) as *a dual of* (IV), as follows:

(VII) *There exist vectors* $x = (x_1, x_2, \cdots)$ *for which the system* $My = x$ *has a solution* $y = (y_1, y_2, \cdots)$ *satisfying*

$$\text{(11 bis)} \qquad \sum_{n=1}^{\infty} | y_n | < \infty,$$

although Möbius' solution Ex, *instead of representing this* y, *does not exist.*

It is clear from (9) and (3 bis) that, if $y = (y_1, y_2, \cdots)$ is any vector for which the series

$$\text{(31)} \qquad \sum_{n=1}^{\infty} y_n$$

is absolutely convergent, then the vector My exists. Let this vector be denoted by x. Then the case $n = 1$ of (8) and the defining relations (9) show that the first component of Ex is the repeated series

$$\text{(32)} \qquad \sum_{n=1}^{\infty} \left(\sum_{k=1}^{\infty} \mu(k) y_{nk} \right).$$

Since y satisfies $My = x$, it follows that (VII) will be proved if it is shown that the absolute convergence of the series (31) does not imply the convergence of the repeated series (32).

This is a precise dual of what has been proved for (22), (23) in **6.** Correspondingly, the following proof parallels that given in **6.** But the proof is by no means superfluous, since (I) and everything that follows (I) prevent a general principle of duality.

Suppose that the series (31) is absolutely convergent, and let t_m denote the m-th partial sum of the exterior summation in (32), that is, let

$$t_m = \sum_{n=1}^{m} \sum_{k=1}^{\infty} \mu(k) y_{nk}.$$

This can be rearranged into

$$t_m = \sum_{j=1}^{\infty} y_j \sum_{nk=j}^{n \leqq m} \mu(k),$$

where the interior summation runs over those positive integers k corresponding to which there exists *some* n not exceeding m and satisfying $nk = j$. Accordingly,

$$t_m = \sum_{j=1}^{\infty} y_j \sum_{d|j}^{d \leqq m} \mu(j/d).$$

Hence, if j is replaced by n,

(33) $$t_m = \sum_{n=1}^{\infty} \phi^n(m) y_n,$$

where $\phi^n(m)$ denotes the interior sum, that is,

(34) $$\phi^n(m) = \sum_{d|n}^{d \leqq m} \mu(n/d).$$

Since (33), $t_m, \phi^n(m)$ correspond to (27), $s_m, \phi_n(m)$ respectively, what corresponds to (18) is

(35) $$\limsup_{n\to\infty,\ m\to\infty} |\phi^n(m)| = \infty.$$

Hence, the proof will be complete if it is shown that (35) is true.

To this end, let the summation index d in (34) be replaced by n/d. Then (34) appears in the form

$$\phi^n(m) = \sum_{d|n}^{d \geqq n/m} \mu(d).$$

Consequently, if n/m is not an integer,

$$\phi^n(m) + \phi_n([n/m]) = \sum_{d|n} \mu(d),$$

by (15). It follows therefore from (3) that (35) is implied by (15) and (18). In fact, the omission of the assumption that n/m is not an integer introduces an error which is bounded, hence such as to have no influence on (35).

9. The content of (IV) is that there are data $(y_1, y_2, \cdots)$ for which the regular solution $(x_1, x_2, \cdots)$ of (8) exists but is not attainable by Möbius' inversion (9). However, the resulting pathological y-range does not contain data $y = (y_1, y_2, \cdots)$ which have occurred thus far in the applications of Möbius' inversion to problems occurring in the analytic theory of numbers. Correspondingly, what really matters in those classical applications

is neither just the existence of a regular solution nor just the existence of a Möbius solution, but the existence of *both.*

Thus there arises the problem of delimiting the y-range within which $\mathrm{E}x = y$ has a regular solution x represented by Möbius' inversion, $\mathrm{M}y$. A regular solution x of this particular type will be called *hyper-regular.* A complete characterization of the y-range of hyper-regularity involves properties of a y which are just as obscure as are, in view of (ii)–(iii) and (i*)–(ii*) in **7**, the properties characteristic of a vector y belonging to the more inclusive range of regularity. However, the following sufficient criterion comprises more than what is needed in the classical applications.

(VIII) *If the data y_n of the system* $\mathrm{E}x = y$, *where* $y = (y_1, y_2, \cdots)$, *satisfy the condition*

$$\sum_{n=1}^{\infty} 2^{\nu(n)} \mid y_n \mid < \infty, \tag{36}$$

then there must exist an (*or, according to* (V), *the*) *hyper-regular solution* x *of* $\mathrm{E}x = y$.

For instance, this will be the case if

$$\sum_{n=1}^{\infty} n^{\epsilon} \mid y_n \mid < \infty \tag{36 bis}$$

holds for some $\epsilon > 0$, since (36 bis) is sufficient for (36), by (29).

In order to prove (VIII), let x denote the vector $\mathrm{M}y$. This vector exists (simply because its components x_n are given by the series (9), the convergence of which is assured by (3 bis) and (36), since (36) implies that (36 bis) is satisfied by $\epsilon = 0$). Furthermore, the components of $x = \mathrm{M}y$ satisfy the regularity conditions (11). In fact, (9) shows that the series (11) becomes the repeated series on the left of the obvious inequality

$$\sum_{n=1}^{\infty} \left| \sum_{k=1}^{\infty} \mu(k) y_{nk} \right| \leqq \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} | \mu(k) | \; | y_{nk} |. \tag{37}$$

But it is clear from (21) that the double series on the right of this inequality, a series of non-negative terms, can be contracted into the simple series (36).

This proves that, if (36) is satisfied, Möbius' vector $\mathrm{M}y$ exists and represents either a regular solution of $\mathrm{E}x = y$ or no solution at all. Hence, in order to complete the proof of (VIII), it suffices to rule out the second of these possibilities.

10. The assertion is that the vector $\mathrm{E}(\mathrm{M}y)$ exists and is identical with the given vector y (provided that (36) is satisfied by the components of y). But substitution of (9) into (8) shows that the assertion $\mathrm{E}(\mathrm{M}y) = y$ can be written in the form

$$(38) \qquad \sum_{m=1}^{\infty}\left(\sum_{k=1}^{\infty}\mu(k)y_{nmk}\right) = y_n, \qquad (n = 1, 2, \cdots),$$

where it is understood that the convergence of the repeated series (38) is part of the statement. Accordingly, the proof of (VIII) will be complete if it is verified that (36) implies the truth of (38) for every n.

First, (36) implies that

$$(39) \qquad \sum_{m=1}^{\infty}\sum_{k=1}^{\infty}|\mu(k)|\,|y_{nmk}| < \infty, \qquad (n = 1, 2, \cdots).$$

In fact, whether (39) be true or false, the double series (39), having only non-negative terms, can be rearranged into

$$\sum_{j=1}^{\infty}|y_j|\sum_{nmk=j}|\mu(k)|,$$

where the interior summation extends over the range of those values k corresponding to which there exists *some* m satisfying the condition $nmk = j$ (in which both n and j are fixed). This means that the double series (39) is identical with

$$(40) \qquad \sum_{n=1}^{\infty}|y_j|\sum_{nk|j}|\mu(k)|,$$

and is therefore majorized by

$$\sum_{j=1}^{\infty}|y_j|\sum_{k|j}|\mu(k)|$$

(simply because $nk|j$ cannot hold for any n if it does not hold for $n = 1$). But (21) shows that this majorant of the double series (39) is identical with the series (36). This proves that (36) implies (39).

Since the repeated series (40) is just a rearrangement of the double series (39), and since the latter is convergent and consists of the absolute values of the terms occurring in the repeated series (38), it follows that the repeated series (38) is convergent and that, in addition, the assertion (38) can be rearranged, corresponding to (40), into

$$\sum_{j=1}^{\infty}y_j\sum_{nk|j}\mu(k) = y_n, \qquad (n = 1, 2, \cdots).$$

Hence, in order to complete the proof of (38), all that remains to be ascertained is that the (finite) sum multiplying y_j on the left of the last formula line is 0 or 1 according as $j \neq n$ or $j = n$. If j and the summation index k are replaced by m and d respectively, this assertion appears in the form

$$(41) \qquad \sum_{nd|m}\mu(d) = e_{nm},$$

where (e_{nm}) denotes the infinite unit matrix. But (41) is true. In fact, if m is not divisible by n, then e_{nm} is 0 (since $n = m$ would imply that m is divisible by m), and the sum on the left of (41) is vacuous (since nd cannot divide m for any d if n itself does not). In the remaining case, that is, if the quotient m/n is an integer, say k, the assignment $nd|m$ on the left of (41) can be replaced by $d|k$, and so the truth of (41) follows from (3) in this case (simply because the integer $k = m/n$ is greater than or equal to 1 according as $m \neq n$ or $m = n$).

11. The criterion (VIII), the proof of which is now complete, places the restriction on the data and is, therefore, an existence theorem. In contrast, the following (incomplete) dual of (VIII) will assume the existence of a solution of a certain restricted type; and all that will be claimed is that the assertion of (VIII) then becomes tautological in some respect.

(IX) *If* $\mathrm{E}x = y$ *has a solution* $x = (x_1, x_2, \cdots)$ *satisfying*

$$\sum_{n=1}^{\infty} 2^{\nu(n)} |x_n| < \infty, \tag{42}$$

then this solution is hyper-regular. In fact, if $\mathrm{E}x = y$ *and* (42) *are satisfied, then* $\mathrm{M}y$ *exists and is precisely* x.

The existence of $\mathrm{M}y$ means, of course, the convergence of the series (8). Hence, the truth of (IX), no matter how elementary, is curious indeed, since the situation is as follows:

(IX bis) *The assumptions of* (IX), *which imply the existence of* $\mathrm{M}y$, *do not imply that*

$$\sum_{n=1}^{\infty} |y_n| < \infty, \tag{43}$$

(*although nothing short of* (43) *appears to guarantee the existence of* $\mathrm{M}y$, *that is, the convergence of all the series* (9), *if* $y = (y_1, y_2, \cdots)$ *is a* free *variable*; *actually, the* y_n *are* bound *by* (43) *and* $\mathrm{E}x = y$).

In order to see this, let $x_1, x_2, \cdots$ be a sequence of positive values satisfying (42) and (30). The possibility of choosing such an $x = (x_1, x_2, \cdots)$ is assured by (4) and (5). Let $y = (y_1, y_2, \cdots)$ be defined by $\mathrm{E}x = y$. This does define the values y_n, since (42) implies (11) and, therefore, the convergence of the series (8). Accordingly, the assumptions of (IX) are satisfied. Nevertheless, (43) fails to hold. This follows from (30), if the restriction $x_n > 0$ is used in the same way as at the end of **7**.

The assertion of (IX) is that (42) and (8) imply (9). But substitution of (8) into (9) gives

$$(44) \qquad \sum_{m=1}^{\infty} \mu(m) \left(\sum_{k=1}^{\infty} x_{nmk}\right) = x_n, \qquad (n = 1, 2, \cdots).$$

Hence, the assertion of (IX) is that (42) implies (44). On the other hand, the proof of (VIII) in **10** consisted in verifying that (36) implies (38). And this depended only on the fact that (36) implies (39). Since (39) *remains unaltered if the summation indices m, k are interchanged,* it follows, on replacing every y_i by the corresponding x_i, that it is superfluous to repeat the details.

12. This formal procedure supplies some, mostly of course superficial criteria relating to the problem which is the dual of Möbius' inversion, namely to the problem of the system $\mathrm{M}y = x$ in which y is the unknown and $\mathrm{E}x$ represents the formal Möbius solution. The simplest fact which can thus be obtained is as follows:

(X) *For any given x, the system $\mathrm{M}y = x$ has at most one solution $y = (y_1, y_2, \cdots)$ satisfying*

$$\sum_{n=1}^{\infty} 2^{\nu(n)} \, | y_n | < \infty.$$

This is a partial dual of (V). A complete dual would not postulate more than

$$\sum_{n=1}^{\infty} | y_n | < \infty,$$

the formal y-analogue of the assumption (11) of (V). That *some* restriction of y is necessary, follows from the fact that the system (13), which is the homogeneous system $\mathrm{M}y = 0$, has a solution y distinct from the trivial solution, $y = 0$.

In order to prove (X), suppose that $\mathrm{M}y = x$ has a solution, $y = y(x)$, satisfying the assumption of (X). This means that both (9) and (36) hold for a certain $y = (y_1, y_2, \cdots)$. But (21) shows that (9) and (36) imply (39). On the other hand, it is clear from (9) that the double series (39) majorizes the series (11). Hence, (11) is satisfied (and so, in particular, $\mathrm{E}x$ exists). Consequently, if n in (9) is replaced by nk, the resulting equations can be summed with respect to k. This leads to the relations

$$(45) \qquad \sum_{k=1}^{\infty} \left(\sum_{m=1}^{\infty} \mu(m) y_{nmk}\right) = \sum_{k=1}^{\infty} x_{nk}, \qquad (n = 1, 2, \cdots).$$

In addition, the repeated series on the left of (45) is the repeated series (38). Since, as verified in **10**, the assumption (36) implies that (38) is an identity, it follows that the repeated series on the left of (45) is identical with y_n. Consequently, (45) means that $y = \mathrm{E}x$. Since $y = \mathrm{E}x$ implies that y is *determined* by x, the proof of (X) is complete.

It is also seen that (X) can be amplified as follows:

(X*) *If* $\mathrm{M}y = x$ *has, for a given vector* x, *a solution* y *satisfying the assumption of* (X), *then this solution must be Möbius' formal solution, that is, the vector* $\mathrm{E}x$, *which exists, since* x *must be a regular solution of* $\mathrm{E}x = y$ *by virtue of the* y*-assumption of* (X).

13. However, (X*) does not supply any criterion discerning between the two cases allowed by (X), the case of non-existence and the case of unique existence. Such a criterion, namely, an existence statement corresponding to a dual of (iv) in (VI), is contained in the fourth of the following assertions:

(X bis) *If* x *is given, and* y *is the unknown, in the system* $\mathrm{M}y = x$, *then*

(i) *the condition*

$$\sum_{n=1}^{\infty} 2^{\nu(n)} \mid x_n \mid < \infty$$

is sufficient in order that Möbius' formal solution, which is $y = \mathrm{E}x$, *should actually be a solution of* $\mathrm{M}y = x$, *but*

(ii) *the condition of* (i) *is insufficient for the existence of a solution* y *for which the restriction*

$$\sum_{n=1}^{\infty} \mid y_n \mid < \infty$$

is satisfied, although

(iii) *the somewhat stricter condition*

$$\sum_{n=1}^{\infty} \tau(n) \mid x_n \mid < \infty$$

is sufficient for the existence of a solution y *satisfying the restriction of* (ii), *and*

(iv) *the still stricter condition*

$$\sum_{n=1}^{\infty} n^{\epsilon} \mid x_n \mid < \infty$$

(*to be satisfied by* some $\epsilon > 0$) *is sufficient for the existence of a solution* y *for which the restriction assumed in* (X) *is fulfilled.*

First, it is clear from the comments made at the end of **11** that, in order to prove (i), it is sufficient to ascertain that the x-assumption of (i) implies the sequence of conditions which results when y is replaced by x in (39). In other words, it is sufficient to ascertain that (36) implies (39). But the truth of this implication was verified after (40).

What concerns (ii), it is enough to take a glance at the proof of the negation in (IX bis).

Correspondingly, (iii) may be verified as follows: According to (4), the x-condition of (iii) implies the x-condition of (i) and so, by the assertion of (i), the existence of a solution $y = (y_1, y_2, \cdots)$ satisfying (8). Consequently, there exists a solution y satisfying

$$\sum_{n=1}^{\infty} | y_n | \leqq \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} | x_{nm} |.$$

Since the double series on the right can be contracted into the series the convergence of which is the x-condition assumed in (iii), the assertion of (iii) follows.

14. It also follows that the x-assumption of (iii) implies the existence of a solution $y = (y_1, y_2, \cdots)$ satisfying the inequality

$$\sum_{n=1}^{\infty} 2^{\nu(n)} | y_n | \leqq \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} 2^{\nu(n)} | x_{nm} |,$$

in which, however, the double series on the right need not converge. But this (non-negative) double series, be it convergent or not, can be rearranged into

$$\sum_{j=1}^{\infty} | x_j | \sum_{nm=j} 2^{\nu(n)},$$

where it is understood that the interior sum denotes $\tau^*(j)$, if τ^* is the arithmetical function defined by

$$\tau^*(n) = \sum_{d|n}^{\infty} 2^{\nu(d)}. \tag{46}$$

If this is inserted into the last inequality, there results the following criterion

(which is *sharp,* since the inequality becomes an equality when every y_n is chosen to be positive; cf. the end of **7**).

(iv bis) *If the data* $x_1, x_2, \cdots$ *of the system* $\mathrm{M}y = x$ *satisfy the condition*

$$\sum_{n=1}^{\infty} \tau^*(n) \mid x_n \mid < \infty,$$

then there exists a solution $y = (y_1, y_2, \cdots)$ *satisfying the assumption of* (X).

This sharp criterion relates to (iv) in the same way as the fourth assertion of (VI) relates to (iv bis), **7**. In other words, (iv) is a corollary, since

$$\tau^*(n) = O(n^\epsilon) \tag{47}$$

holds for every $\epsilon > 0$. In fact, since the logarithm of

$$\tau(n) = \sum_{d|n} 1 \tag{48}$$

is subject to the well-known estimate

$$\log \tau(n) = O(\log n)/\log \log n \tag{28 bis}$$

(Wigert-Ramanujan), the estimate

$$\tau(n) = O(n^\epsilon) \tag{29 bis}$$

holds for every $\epsilon > 0$. Hence, (4) and (48) imply the estimate

$$\sum_{d|n} 2^{\nu(d)} \leqq \sum_{d|n} \tau(d) = O(n^\epsilon) \sum_{d|n} 1 = O(n^{2\epsilon}),$$

which, in view of (46), is the assertion (47).

Needless to say, the necessity of replacing the condition of (iii) by the condition of (iv bis) is due to the fact that

$$1 \leqq \tau(n) \leqq \tau^*(n) \tag{4 bis}$$

but

$$\limsup_{n\to\infty} \tau^*(n)/\tau(n) = \infty. \tag{5 bis}$$

This is clear from (46) and (48), since $2^{\nu(n)} \geqq 1$ but $\limsup_{n\to\infty} 2^{\nu(n)} = \infty$.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] H. Bruns, *Die Grundlinien des wissenschaftlichen Rechnens*, Leipzig, 1903.

[2] A. Haar, Chapter I, no. 129 in G. Szegö and G. Pólya, *Aufgaben und Lehrsätze aus der Analysis*, Berlin, vol. 1 (1925).

[3] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford, 1938.

[4] E. Hellinger and O. Toeplitz, "Grundlagen für eine Theorie der unendlichen Matrizen," *Mathematische Annalen*, vol. 69 (1910), pp. 289-330.

[5] E. Helly, "Ueber Systeme linearer Gleichungen mit unendlich vielen Unbekannten," *Monatshefte für Mathematik und Physik*, vol. 31 (1921), pp. 60-91.

[6] A. Rajchman, "Ueber eine paradoxe Eigenschaft gewisser bedingt konvergenter unendlicher Reihen," *Mathematische Zeitschrift*, vol. 26 (1927), pp. 777-778.

[7] I. Schur, "Ueber lineare Transformationen in der Theorie der unendlichen Reihen," *Journal für die reine und angewandte Mathematik*, vol. 151 (1920), pp. 79-111.

[8] A. Wintner, *An Arithmetical Approach to Ordinary Fourier Series*, Baltimore, 1945.

METRICALLY HOMOGENEOUS SPACES.*

By Herbert Busemann.

The following note shows how certain questions of metric geometry relating to the so-called four-point properties [1] can be unified and derived from a general theorem which is in turn a simple consequence of a geometric result obtained elsewhere.[2] The theorem states essentially this: a metric space R with suitable compactness and convexity properties has constant curvature when for any linear triple of points q, r, x and any fourth point p the distance px is a function $\phi(pq, pr, xq, \pm xr)$. Thus it will be *proved* that ϕ is one of three specific functions (see formulas (1), (2), (3)) occurring in euclidean, hyperbolic, or spherical geometry respectively, whereas any of the four point properties *assumes* a priori that ϕ is one of these specific functions.

The exact formulation of the theorem is this:

Theorem. *The space R is a locally isometric map of a finite dimensional euclidean, hyperbolic, or spherical space* [3] *if and only if it satisfies the following five conditions:*

I. *R is metric (with distance xy).*
II. *R is finitely compact.*
III. *R is convex.*
IV. *R is locally externally convex.*

If $S(z, \rho)$ denotes the set of points with $zx < \rho$, then IV means more explictly this: for every point a there is a $\rho_a^1 > 0$ such that for any two points p, q in $S(a, \rho_a^1)$ a point $r \neq q$ with $pq + qr = pr$ exists.

V. *For every point a there is a $\rho_a^2 > 0$ and a function $\phi_a(\xi_1, \xi_2, \xi_3, \xi_4)$ such that for any four points p, q, r, x in $S(a, \rho_a^2)$ with $qx + \epsilon xr = qr > 0$, $\epsilon = \pm 1$, the relation*

* Received May 23, 1945.

[1] For the literature see Blumenthal [1], in particular pp. 69-84.

[2] The result referred to states that spaces with locally linear bisectors have constant curvature, see Busemann [1] (quoted as B.) p. 268.—The author takes this occasion to point out that Theorem (1.21) in B, which characterizes the Hausdorff spaces with a finitely compact metrization, is not new, but was proved previously by H. E. Vaughan, see Vaughan [1], p. 532, Theorem 2.

[3] More briefly we shall say that R has constant curvature.

$$px = \phi_a(pq, pr, xq, \epsilon xr)$$

holds.

The theorem would not be true if V were required for $\epsilon = 1$ only, even if IV is replaced by external convexity in the large. An example is furnished by the space consisting of the three rays $0 \leq \tau^i < \infty$ $(i = 1, 2, 3)$, with the metric

$$\tau_1{}^i \tau_2{}^j = \begin{cases} |\tau_1{}^i - \tau_2{}^j| & \text{if } i = j, \\ \tau_1{}^i + \tau_2{}^j & \text{if } i \neq j. \end{cases}$$

This definition implies that the three origins $\tau^i = 0$ are identified. It is easily seen that IV holds in the large and that

$$px = \max(pq - xq, pr - xr)$$

whenever $qx + xr = qr$, so that V holds in the large for $\epsilon = 1$. But no neighborhood of $\tau^i = 0$ is even homeomorphic to the interior of a sphere of any E^n.

The necessity of the hypotheses I to V is obvious because the Pythagorean theorems of the geometries with constant curvatures 0, $-\beta^2$, β^2 yield respectively,

(1) $px^2 = [pq^2 \cdot \epsilon \cdot xr + pr^2 \cdot xq - xq \cdot \epsilon \cdot xr \cdot qr)/qr]$

(2) $\cosh(\beta px) = [\cosh(\beta pq)\sinh(\epsilon\beta xr) + \cosh(\beta pr)\sinh(\beta xq)]/\sinh(\beta qr)$

(3) $\cos(\beta px) = [\cos(\beta pq)\sin(\epsilon\beta xr) + \cos(\beta pr)\sin(\beta xq)]/\sin(\beta qr)$.

To see the sufficiency put $\delta_a{}^i = \sup \rho_a{}^i$ $(i = 1, 2)$, where $\rho_a{}^i$ traverses all numbers for which $S(a, \rho_a{}^i)$ satisfies IV or V respectively. If $b \in S(a, \delta_a{}^i)$ then $S(b, \delta_a{}^i - ab) \subset S(a, \delta_a{}^i)$, so that $\delta_b{}^i \geq \delta_a{}^i - ab$; hence by symmetry

(4) $$|\delta_a{}^i - \delta_b{}^i| \leq ab, \qquad (i = 1, 2).$$

Because of I, II, III any two points x, y of R can be connected by a segment $\mathfrak{t}(x, y)$. If the three points b, c, d are different and $bc + cd = bd$ we write (bcd). The relations (cbe) and (bde) imply (cbd) and (cde) and conversely (compare B, Section 1). The following is a consequence of conditions I to IV.

(5) If $p, q \in S(a, \delta_a{}^1/3)$ then any segment $\mathfrak{t}' = \mathfrak{t}(p, q)$ is a subsegment of a segment $\mathfrak{t}(p', q')$ with $ap' = aq' = 2\delta_a{}^1/3$.

For by the preceding remarks and IV there are pairs of points x', y' with

$ax' \leq 2\delta_a/3$, $ay' \leq 2\delta_a^1/3$, (pqy'), $(x'py')$ hence also $(x'pq)$ and $(x'qy')$. Because of II these pairs x', y' form with the natural metric (that is $(x'_1, y'_1)(x'_2, y'_2) = x'_1x'_2 + y'_1y'_2$) a non-empty compact set; hence there is a pair p', q' for which $x'y'$ reaches its maximum $p'q'$. Then p', q' satisfy (5). For if $ap' < 2\delta_a^1/3$ a point x^* with $(x^*p'q')$ and $ax^* \leq 2\delta_a^1/3$ would exist. But then also (x^*pq') and $(x^*p'q')$ so that x^*, q' would be an admissible pair x', y' with $x^*q' > p'q'$. Then $\mathfrak{t}(p', p) \cup \mathfrak{t}' \cup \mathfrak{t}(q, q')$ is a segment which satisfies (5).

Next observe the following consequence of I and V

(6) If the points q, r, x_1, x_2 of $S(a, \delta_a^2)$ satisfy the relations $qx_i + \epsilon x_i r = qr > 0$ $(i = 1, 2)$ and $x_1 r = x_2 r$, then $x_1 = x_2$.

For then also $qx_1 = qx_2$, hence

$$x_1x_2 = \phi_a(x_1q, x_1r, x_2q, \epsilon x_2r) = \phi_a(x_2q, x_2r, x_2q, \epsilon x_2r) = x_2x_2 = 0.$$

Now put $\rho_a = \min(\delta_a^1/3, \delta_a^2)$. Then (5) and (6) show that for any two points p, q of $S(a, \rho_a)$ and every $\alpha > 0$ points r in $S(a, \rho_a)$ with (pqr) and $qr < \alpha$ exist, and that (pqr') and $qr = qr'$ imply $r = r'$. Hence the basic axiom *D* of *B*. p. 215 holds, so that *R* is a *G*-space (compare *B*. p. 227).

The only one-dimensional *G*-spaces are the straight line and the great circles (B. p. 233). In this case the following considerations are trivial, therefore the space will be assumed to have at least dimension 2.

If $p, p' \in S(a, \rho_a/2)$ and $p \neq p'$ call $B(p, p')$ the locus of those points x for which $px = px'$. If q, r are points of $B(p, p') \cap S(a, \rho_a/2)$, then a segment $\mathfrak{t}(q, r)$ lies in $S(a, \rho_a)$ (B. 1.15). If x is a point of this segment then

$$px = \phi_a(pq, pr, xq, xr) = \phi_a(p'q, p'r, xq, xr) = p'x$$

so that $\mathfrak{t}(q, r) \subset B(p, p')$. The neighborhood $S(a, \rho_a/2)$ has, therefore, linear bisectors (compare B. p. 262 condition (*)), so that the theorem follows from the First Characterization of the spaces with constant curvature in B. p. 268.

If in any metric space four points p, q, x, r with $qx + \epsilon xr = qr > 0$ are congruent to four points of a hyperbolic space of curvature $-\beta^2$, then the relation (2) holds. A similar remark applies to the euclidean case and to the spherical case if $\beta qr < \pi$. Hence we find

COROLLARY 1. *If R satisfies conditions I to IV and if every point a of R has a neighborhood $S(a, \rho_a^2)$ such that any four points p, q, r, x in $S(a, \rho_a^2)$ with (qxr) are congruent to a quadruple of points in a space $R(a)$ which is euclidean, hyperbolic or spherical, then R has constant curvuture.*

It is of interest to study the consequences of V if required in the large.

COROLLARY 2. *If R satisfies I to IV and if a function $\phi(\xi_1, \xi_2, \xi_3, \xi_4)$ exists such that*

$$px = \phi(pq, pr, xq, \epsilon xr)$$

when $qx + \epsilon xr = qr > 0$, then R is euclidean or hyperbolic.

The Theorem shows that R has constant curvature, whereas (6) implies that the segment $\mathfrak{t}(a, b)$ is unique for any a, b. Any geodesic in R is contained in a two-dimensional surface of constant curvature. The elliptic plane and the sphere are the only surfaces of constant positive curvature;[4] hence every space of constant positive curvature contains points a, b for which $\mathfrak{t}(a, b)$ is not unique.

The only surfaces of non-positive constant curvature on which shortest connections are unique are the euclidean and hyperbolic planes. This proves Corollary 2.

The statement in the large corresponding to Corollary 1 is this:

COROLLARY 3. *If I to IV hold and any quadruple of points in R is congruent to a quadruple of points in a euclidean space or a hyperbolic space of curvature $-\beta^2$ or a spherical space of curvature β^2, then R is a euclidean, hyperbolic or spherical space.*

The formulas (1) and (2) hold in the first two cases, hence the assumption of Corollary 2 are satisfied.

In the spherical case (3) holds only for $\beta qr < \pi$. But since (3) holds in the small the Theorem shows that R has constant positive curvature. Again,[4] the only two-dimensional totally geodesic subspaces of R are spheres and elliptic planes of curvature β^2. An elliptic plane contains four points a_1, a_2, a_3, a_4 (on a geodesic such that $a_1a_2 = a_2a_3 = a_3a_4 = a_4a_1 = \pi\beta/4$, $(a_1a_2a_3)$, $(a_2a_3a_4)$, and $(a_3a_4a_1)$. This quadruple is not congruent to a quadruple on a sphere of radius $1/\beta$.

Condition II in Corollary 3 is stronger than completeness and separability assumed by Wilson and Blumenthal, consequently their results include infinite dimensional spaces. On the other hand in the euclidean and hyperbolic cases, these authors require external convexity, that is IV in the large. (Compare Blumenthal [1], p. 69 Theorem 6. 4).

---

[4] See Cartan [1], p. 174.

Spherical spaces are not externally convex. Therefore Blumenthal [1, p. 74, Theorem 8.3] replaces external convexity by diametrization, which means that for every point p a point p' with $pp' = \pi/\beta$ exists. The present condition IV has the advantage of applying to all three cases.

SMITH COLLEGE.
NORTHAMPTON, MASS.

REFERENCES.

Blumenthal, L. M. [1], *Distance Geometries, University of Missouri Studies* XIII (1938), No. 2.

Busemann, H. [1], "Local metric geometry," *Transactions of the American Mathematical Society*, vol. 56 (1944), pp. 200-274.

Cartan, E. [1], *Leçons sur la géométrie des espaces de Riemann*, Paris, 1928.

Vaughan, H. E. [1], "On locally compact metrizable spaces," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 532-535.

AN EXTENSION OF KLEIN'S ERLANGER PROGRAM: LOGIC AS INVARIANT-THEORY.*

By F. I. Mautner.

Introduction. It will be shown in this paper that two-valued (Boolean) mathematical logic of propositions and propositional functions (in extension) can be considered as *"invariant-theory of the symmetric group"* in the sense of Klein's Erlanger program [1] (Chapter I). Consequently an attempt is made to create the means necessary for developing it as such a "theory of notions of invariant significance and their invariant properties and relations" (Chapters II, III and IV).

The attitude taken towards logic will be dominated by a strong analogy with coordinate-geometry and its invariant-theoretic foundations. Indeed, if taken in extension, a propositional function is determined by its truth-values Boolean 0 and 1, just as a vector or tensor in geometry is determined by its coordinates. And just as the geometrical properties are shown to be independent of any particular choice of the coordinate system by proving their invariance with respect to the group of coordinate-transformations, so the logical notions and properties will be shown to be independent of any particular assignment of truth-values (the "logical coordinate system") by showing their invariance with respect to the "group of logical coördinate-transformations," which will be seen to be the symmetric group of all permutations of the domain of individual variables (assuming only one such domain).

In this connection it ought to be remembered that the group- and invariant-theoretic definition of a geometry has so far been independent of the rigorous axiomatic foundations of geometry. Accordingly two-valued logic will in the following not be based on any particular system of axioms, but taken in its "naive" form as a purely algebraic system. For the aim of this paper is not a contribution to the rigorous foundations of logic, but the application of group- and invariant-theoretic methods. And it is hoped that this will result in an invariantive calculus whose symbolism is as algebraically suggestive and easily handled as the corresponding invariantive calculus for geometry, namely tensor calculus.

---

* Received April 30, 1945; Revised December 15, 1945.

[1] Cf. Weyl's formulation of Klein's Erlanger program (Classical Groups, pp. 13-18) on which much of the following is based.

In order to show that two-valued logic is invariant-theory of the symmetric group we prove first that the group of automorphisms of the calculus of those propositional functions which are defined for every member of one domain C of individuals is the symmetric group $\mathfrak{S}_n$ of all permutations of the (finite or transfinite) domain C. The second fundamental characteristic of an "invariant-theory" is the existence of coordinates in which the group of automorphisms induces an isomorphic group of admissible coordinate-transformations. If one defines any particular assignment of truth-values to the propositions and propositional functions to be a "logical coordinate system," then the symmetric group $\mathfrak{S}_n$ becomes the "group of logical coordinate-transformations" with respect to which the logical notions and their properties are invariant. This can be given an exact meaning by showing that Weyl's axioms for Klein's Erlanger program (*loc. cit.*) are satisfied. Hence under the assumption of only one domain of individuals we have the result that only such properties of n^r-tuples of Boolean 0's and 1's ($n =$ cardinal number of individuals, $r = 0, 1, 2, 3, \cdots$), are of objective logical meaning as are invariant under the symmetric group.

This result is very analogous to the well known state of affairs in geometry: Each linear geometry [2] is the invariant-theory of a group of linear transformations and could be systematically characterised as such. This suggests the same for logic: Firstly an appropriate tensor algebra (Chapter II), secondly a theory of the possible transformation laws (Chapter III) and thirdly a theory of the invariant properties of objects transforming according to these possible transformation laws (Chapter IV).

The appropriate tensor algebra is obtained by observing that there is a strong analogy between a propositional function of r arguments (and the operations of the calculus of propositional functions) and a tensor of rank r (and the operations of tensor algebra). One will therefore define a "*Boolean tensor of rank r*" to be an object which is, relative to any fixed logical coordinate system, defined by an n^r-tuple of Boolean 0's and 1's, with the transformation-law "r-fold Kronecker product" of $\mathfrak{S}_n$ with itself (Kronecker product in the sense of matrices). Then one obtains what I call "*Boolean tensor algebra*" which is isomorphic to the calculus of propositional functions (over *one* domain of individuals) as far as conjunction, disjunction, implication, negation and quantification of propositional functions are concerned. Boolean tensor algebra, though very analogous to ordinary tensor algebra, differs from it in essentially three respects: The underlying group is the

[2] We confine our analogy here to the linear geometries; the analogy between other geometries and logic seems much less deep and not so suggestive.

symmetric group, the values of the "components" of a Boolean tensor are Boolean 0 or 1 and there may be transfinitely many components. Owing to the different group there are two contractions here: Sum and product over one index are already invariant; they correspond to the two quantifiers over individual variables.

In order to find all possible transformation laws one has to study appropriate representations of the group of automorphisms. Whereas in geometry one confines oneself to matrix-representations, clearly *any* representation by a group of arbitrary transformations (a "group-realisation") is a possible transformation law here. Some properties of matrix-representations can be deduced from properties of groups with endomorphisms (via the introduction of a "representation module") in virtue of the fact that every matrix is an endomorphism of a vector-space. Similarly the fact that (as will be seen) every permutation is an automorphism of a Boolean algebra leads via the introduction of a realisation-module (= a "Boolean ring with endomorphisms") to analogous theorems on group-realisations (by means of arbitrary permutations). The analogy can be followed up further: all transitive (permutation-) realisations of any group can be obtained from its regular realisation and all transitive realisations of the (finite) symmetric group are contained in Boolean tensor space.

Next one has to define a "kind of *quantity*" to correspond to every possible transformation-law (i. e. to every possible realisation). Just as Weyl's and van der Waerden's quantities for geometry and the theory of invariants form vector spaces with a matrix-representation of the group of automorphisms g induced in them (i. e. representation-modules), so "*Boolean quantities*" form Boolean algebras with a realisation of g induced in them (i. e. realisation-modules). One can thus observe step by step *a close parallelism between the full linear group and affine or projective geometry on the one hand and the symmetric group and Boolean algebra of logic on the other.* And this close analogy reaffirms that not only can logic be considered as invariant-theory of the symmetric group, but also that group- and invariant-theoretic methods should be as fruitful here as in geometry.

Naturally one will now ask for the invariant properties of Boolean quantities. This requires the definition of an invariant property or relation and leads to the definition of *Boolean invariants* and *covariants* which are strictly analogous to the ordinary in- and covariants (in the modern sense). It is shown that the components of a Boolean covariant of any Boolean quantity q are Boolean polynomials of the components of q, no matter what the group of automorphisms. Hence there exists always a basis for the Boolean in-

variants of any Boolean quantity q such that every Boolean invariant is a Boolean polynomial of these "basic invariants," which are themselves Boolean polynomials of the components of q. This basis cannot however always be finite. Finally it is shown that the Boolean invariants (under $\mathfrak{S}_n$) of propositional functions over one finite domain of individuals can all be obtained by the processes of the calculus of propositional functions; in particular the Boolean invariants of one propositional function of one argument can be expressed by numerical conditions.

In conclusion the question as to the possibility of a further extension of Klein's Erlanger program is raised. It is indicated in what way Boolean invariants under the symmetric group might perhaps be applied to yield a classification of formal axiomatic theories. For it is characteristic of the axiomatic method to assume meaningless, hence in particular indistinguishable, individuals. Therefore, if there is only one domain of individuals, the formalism of any completely formal axiomatic theory must be invariant under all permutations of its individuals.

I should like to thank Professor Hermann Weyl for his kind assistance in obtaining very helpful criticism of the original manuscript which led to its revision.

CHAPTER I.

Invariance under the Symmetric Group.

1. The group of automorphisms of the calculus of propositional functions. We assume *one domain C of individuals* to be given as *a priori fixed,* and that *every propositional function $f(x)$ is defined for every x in C.* Every $f(x)$ is then fully determined by that subset C_1 of C such that $f(x)$ is true if and only if x is in C_1. Hence the Boolean algebra of all $f(x)$ is isomorphic to the Boolean algebra of all subsets C_1 of C (called the "full Boolean algebra of dimension n" where n is the cardinal of C).

THEOREM 1.1. *The group of automorphisms of the Boolean algebra B_n of all subsets of a set C is the symmetric group $\mathfrak{S}_n$ of all permutations of C.* (Proof obvious).

One now sees the necessity of an *a priori fixed* domain C of individual variables: Otherwise one could not speak of *the group* of automorphisms.

It follows incidentally that every automorphism of a finite Boolean algebra B_n induces and is induced by an automorphism of the lattice of all Boolean subalgebras

of B_n. For every finite Boolean algebra is full and G. Birkhoff[3] has shown that the lattice of all Boolean subalgebras is dually isomorphic to the lattice $E(C)$ of all equivalence relations over C if n is finite and that the group of automorphisms of $E(C)$ is the symmetric group $\mathfrak{S}_n$ on C.[4]

Thus the group of automorphisms of the calculus of propositional functions of one variable is—with respect to conjunction, disjunction, implication and negation—the symmetric group $\mathfrak{S}_n$. A propositional function $f(x, y)$ is determined by that subset of the totality of ordered pairs (x, y) with x and y in C, for which it takes the value 1 (true). I. e., we make the *assumption* that *the domain and converse domain of every binary relation coincide and is the same for any two* $f(x, y)$, *namely* C. Similarly *all propositional functions* $f(x_1, \cdots, x_r)$ *of* r *arguments under consideration are assumed to have the same range of definition, namely the set of all ordered* r*-tuples of elements of* C. Under this restriction the propositional functions of r arguments form with respect to conjunction, disjunction, implication and negation a Boolean algebra $[B_r]_r$ isomorphic to the Boolean algebra of all subsets of the set $[C]_r$ of all ordered r-tuples of elements of C.[5] The permutations of C induce in $[C]_r$ a subgroup of the symmetric group of all permutations of $[C]_r$, every one of which is by Theorem 1.1 an automorphism of $[B_n]_r$. Hence $\mathfrak{S}_n$ induces in $[C]_r$ a subgroup of the group of automorphisms of $[B_n]_r$.

That quantifiers such as

$$\sum_{x_{i_1}} \prod_{x_{i_2}} \cdots \sum_{x_{i_k}} f(x_1, x_2, \cdots, x_r)$$

are invariant follows at once: Each $\sum_{x_i}$ or $\prod_{x_i}$ is a symmetric function of all the values of f as x_i ranges over C and hence it is invariant.

Again the two quantifiers

$$\sum_{f \in [B_n]_r} f(x_1, \cdots, x_r), \qquad \prod_{f \in [B_n]_r} f(x_1, \cdots, x_r)$$

ranging over all propositional functions of r arguments are clearly invariant.

Besides conjunction and disjunction of propositional functions of the same

---

[3] G. Birkhoff, *Lattice Theory* (New York 1940), Theorem 6.7. We adopt the terminology of this book.

[4] G. Birkhoff, *Proceedings of the Cambridge Philosophical Society*, vol. 31, p. 449.

[5] J. C. C. McKinsey has given an axiomatisation of the calculus of binary relations, every realisation of which is isomorphic to the Boolean algebra of the set of all ordered couples from some domain of individuals. (*Journal of Symbolic Logic*, vol. 5 (1940), p. 94, Theorem B).

variables, there are conjunctions and disjunctions of propositional functions of different variables. E. g.,

$$f(x) \,\&\, g(y) = h(x, y); \qquad f(x) \vee g(y) = k(x, y).$$

Let S be any permutation of C and denote by $S \cdot f(x)$, $S \cdot h(x, y), \cdots$ the propositional functions into which $f(x)$, $h(x, y), \cdots$ are transformed by S. I. e., let $S \cdot f(x) = f(Sx)$, $S \cdot h(x, y) = h(Sx, Sy)$ where $Sx, Sy, \cdots$ are the elements of C into which $x, y, \cdots$ are transformed by S. Then

$$\begin{aligned} & S \cdot (f(x) \,\&\, g(y)) = S \cdot h(x, y) = h(Sx, Sy) \\ \text{and} \quad & S \cdot f(x) \,\&\, S \cdot g(y) = f(Sx) \,\&\, g(Sy) = h(Sx, Sy) \\ \therefore \quad & S \cdot (f(x) \,\&\, g(y)) = S \cdot f(x) \,\&\, S \cdot g(y). \end{aligned}$$

Similarly

$$S \cdot (f(x) \vee g(y)) = S \cdot f(x) \vee S \cdot g(y).$$

In exactly the same way it follows that

$$S \cdot (f(x_1, \cdots, x_r) \,\&\, g(x_{r+1}, \cdots, x_t)) = S \cdot f(x_1, \cdots, x_r) \,\&\, Sg(x_{r+1}, \cdots, x_t)$$

and

$$S \cdot (f(x_1, \cdots, x_r) \vee g(x_{r+1}, \cdots, x_t)) = S \cdot f(x_1, \cdots, x_r) \vee S \cdot g(x_{r+1}, \cdots, x_t)$$

i. e. every S in $\mathfrak{S}_n$ is an automorphism also with respect to conjunction and disjunction of propositional functions of different variables.

In order to show now that such formation as the relative sum and relative product

$$\sum_{y \in C} f(x, y) \,\&\, g(y, z), \qquad \prod_{y \in C} f(x, y) \vee g(y, z)$$

of two binary relations $f(x, y)$, $g(x, y)$ are invariant, it only remains to remark that to put equal variables in a propositional function is clearly invariant under permutations of C. This completes the proof of

THEOREM 1.2. *The group of automorphisms of the calculus of propositional functions is—under the assumption of one a priori fixed domain C of individual variables—the symmetric group $\mathfrak{S}_n$ of all permutations of C, provided every $f(x_1, x_2, \cdots, x_r)$ is uniquely defined for every x_k in C and no other x.*[6]

---

[6] The following remark is due to one of the referees: The extended group of automorphisms may be obtained by interchanging 0 and 1, i. e. by adjoining the fundamental duality of the calculus of propositional functions to its group of automorphisms, exactly as one adjoins the polarities to the group of collineations.

2. Logical coordinates; principle of relativity. In order to be able to give an exact meaning to "independence of the particular assignment of truth-values" we shall now define logical coordinates in strict analogy to the coordinates of geometry:

Suppose the system Φ of all propositional functions of r arguments x in C to be abstractly given, without any particular assignment of truth-values, for instance by formal axioms. Next introduce the system of all possible truth-functions or "logical coordinates" consisting of all n^r-tuples ξ of Boolean 0 and 1, where 0 and 1 occur together n^r times in ξ ($n =$ cardinal of C).

Any isomorphism between the system Φ of abstractly given propositional functions f of r arguments and the system Ξ of all n^r-tuples ξ of Boolean 0 and 1 $(r = 0, 1, 2, \cdots)$ is defined to be an admissible logical coordinate-system. ($r = 0$ includes atomic propositions as a special case).

There are as many isomorphisms between Φ and Ξ as there are automorphisms of Φ (or Ξ),[7] hence $n!$ different equally admissible logical coordinate-systems,[8] in the sense that if any one coordinatisation

$$f \leftrightarrow \xi$$

is given, every other equally admissible one is obtained by applying an automorphism to Φ (or Ξ). Hence the group of transitions between equally admissible logical coordinate-systems is induced by the group of automorphisms of Φ, i. e. by the symmetric group $\mathfrak{S}_n$ on C. It is the realisation of $\mathfrak{S}_n$ as the group of regular permutations of the class of all $n!$ equally admissible logical coordinate-systems (the "regular realisation of $\mathfrak{S}_n$").

An alternative way of introducing logical coordinates is the following: Again suppose the system Φ of propositional functions to be abstractly given. Let Φ_1 be the set of all propositional functions $f_k(x)$ of one argument which are true for one and only one x $(= x_k)$. [To be true for a given number of individuals is of abstract objective meaning, since it is invariant]. Then any other $f(x)$ is the unique logical sum of all the $f_k(x)$:

$$f(x) = \sum_{k \in K} \epsilon_k \cdot f_k(x) \text{ with } \epsilon_k = \begin{cases} 1 \text{ if } f(x) \supseteq f_k(x)^{9} \\ 0 \text{ if } f(x) \not\supseteq f_k(x). \end{cases}$$

Similarly every $f(x_1, \cdots, x_r)$ can be expressed uniquely in the form

$$f(x_1, x_2, \cdots, x_r) = \sum_{k_1, k_2, \ldots, k_r \in K} \epsilon_{k_1 k_2 \ldots k_r} \cdot f_{k_1}(x_1) \cdot f_{k_2}(x_2) \cdots f_{k_r}(x_r)$$

[7] This is true for any two isomorphic algebraic systems.

[8] By $n!$ we mean the cardinal number of all permutations of n things.

[9] $f \supseteq g$ means g implies f, $f \cdot g$ means logical product of f and g.

where

$$\epsilon_{k_1k_2\ldots k_r} = \begin{cases} 1 \text{ if } f(x_1,\cdots,x_r) \supseteq f_{k_1}(x_1)\cdot f_{k_2}(x_2)\cdots f_{k_r}(x_r) \\ 0 \text{ otherwise} \end{cases} \quad \cdots (1)$$

determines the ϵ's uniquely and where K is an arbitrary index set with the same cardinal n as C.

This uniquely determined set of $\epsilon_{k_1\ldots k_r} = 1$ or 0 can be taken as the logical coordinates of $f(x_1, x_2, \cdots, x_r)$. Suppose $\epsilon^{(1)}{}_{k_1\ldots k_r}$, $\epsilon^{(2)}{}_{k_1\ldots k_r}$ are the logical coordinates of $f^{(1)}(x_1, x_2, \cdots, x_r)$, $f^{(2)}(x_1, \cdots, x_r)$ respectively, obtained from equation (1). Clearly $\epsilon^{(1)}{}_{k_1\ldots k_r} = \epsilon^{(2)}{}_{k_1\ldots k_r}$ for all $k_1, \cdots, k_r$ in K if and only if $f^{(1)}(x_1, \cdots, x_r) = f^{(2)}(x_1, \cdots, x_r)$ for all $x_1, \cdots, x_r$ in C.

Hence the correspondence

$$f(x_1, \cdots, x_r) \leftrightarrow \epsilon_{k_1\ldots k_r}$$

defined by equation (1) is one-one. Moreover it is easily seen to be an isomorphism with respect to the operations of the calculus of propositional functions.

Every one of the $n!$ permutations of C induces a unique permutation of Φ_1 and hence a transition to another one-one correspondence $f \leftrightarrow \epsilon$. Clearly the class of $n!$ coordinatisations $f \leftrightarrow \epsilon$ obtained in this way is equal to the above class of $n!$ equally admissible logical coordinate-systems $f \leftrightarrow \xi$. It follows in particular that any logical coordinate-system is determined when the coordinates of all those propositional functions $f(x)$ of one argument are given which are true for one and only one value of x. *Hence any two one-one correspondences between the set Φ_1 of all abstractly given propositional functions which are true for exactly one x and the set Ξ_1 of all n-tuples of* 0 *and* 1 *with exactly one* 1 *determine equally admissible logical coordinate-systems and they are the only ones.*

The specification of one such logical coordinate-system is arbitrary; at best we can determine *objectively the class of equally admissible logical coordinate-systems and ascertain the group of transitions between them.* This is the "relativity-problem" for logic. Its answer is

THEOREM 1. 3 (*Principle of Relativity for Logic*). *The group of transitions between equally admissible logical coordinate-systems is induced by the symmetric group $\mathfrak{S}_n$ of all permutations of the domain C of individuals; it is the regular realisation of $\mathfrak{S}_n$.*

3. Invariant-theory of the symmetric group. The existence of a group of automorphisms and of coordinates, hence of a "principle of relativity" are

the very conditions on the ground of which one can call a geometry an "invariant-theory." Since these conditions are fulfilled here just as in geometry, it results that two-valued mathematical logic is also an invariant-theory in the sense of Klein's Erlanger program. This can, however, be given an exact meaning by *defining a mathematical theory to be an Invariant-Theory of a group g if it satisfies Weyl's axioms for Klein's Erlanger program*[10] *with g as group of automorphisms.*

We then have

THEOREM 1.4. *Two-valued (Boolean) mathematical logic of propositions and propositional functions over one domain C of individual variables is invariant-theory of the symmetric group $\mathfrak{S}_n$ of all permutations of C, provided every propositional function $f(x_1, \cdots, x_r)$ is defined for all x_k in C $(k = 1, 2, \cdots, r)$ and only such x_k.*

Proof. We have to show that Weyl's axioms are satisfied:

A. (1). (Existence of a group g). This is the symmetric group $\mathfrak{S}_n$.

A. (2). (Existence of a set of elements called coordinates and a realisation of g by means of one-one correspondences within that set). The logical coordinates as defined in I. 2, and the regular realisation of $\mathfrak{S}_n$ as group of logical coordinate-transformations.

B. (1). [i]: (Existence of "frames" any two of which determine a group element). Logical coordinate-systems (i. e. isomorphisms $\Phi \leftrightarrow \Xi$) are the required frames, any two of which determine an automorphism of Φ, i. e., an element of $\mathfrak{S}_n$.

[ii] (Every group element shall induce a transition from one frame to another). This has been shown in **2**.

[iii] (The group of transitions is homomorphic to g). This holds by Theorem 1. 3.

B. (2). [i] (Requirement that the objects of the theory be quantities q, i. e., relative to an arbitrarily fixed frame there is a one-one mapping of the possible values of q onto the set of coordinates). By the definition of logical coordinates, as f varies over all propositional functions of r arguments, the coordinates ξ of f vary over all possible n^r-tuples of Boolean 0 and 1 in a one-

[10] Classical Groups, *loc. cit.* or *Duke Mathematical Journal*, vol. 5 (1939), p. 491.

one manner. In particular every atomic proposition has either 0 or 1 as its coordinate.

[ii] (Under transitions to other frames the coordinates of q shall be transformed according to a realisation[11] of g). For a propositional function of r arguments this realisation is the group induced by $\mathfrak{S}_n$ in the set of all ordered r-tuples of elements of C. For an atomic proposition it is the identity-realisation of $\mathfrak{S}_n$, since the two-element Boolean algebra B_1 has no automorphism besides the identity. (In this sense atomic propositions are analogous to scalars). Thus *propositions and propositional functions* (over one domain C of individuals) *satisfy the requirements of the objects of an invariant-theory, i. e. of "quantities."*

This completes the proof.

It is an interesting consequence of the fact that the one-dimensional Boolean algebra B_1 has no automorphisms other than the identity, that to "assert a proposition," i. e., to state that an atomic proposition has the value 1, is of invariant meaning.

4. Several domains of individual variables. Suppose now that there are several domains $C_1, C_2, \cdots, C_p$ of individuals ($p =$ finite). They may be "dependent" or "independent," i. e., if x_j is in C_j and x_k in C_k, x_k may be a function of x_j or not. In two special cases one can still speak of invariance with respect to one or several symmetric groups:

1) Let $C_2, C_3, \cdots, C_p$ be *completely dependent on* C_1, by which we mean that every element x_k in C_k ($k = 2, 3, \cdots, p$) is *a priori* given as such a function of elements x_1 and/or subsets C_1 of C_1 that every permutation of C_1 induces a permutation of C_k. Then it follows at once by the same argument as before that *the group of automorphisms of the calculus of all those propositional functions which are defined for every element of a finite number of a priori fixed domains $C_1, C_2, \cdots, C_p$ of individuals is the symmetric group of all permutations of C_1, provided that $C_2, C_3, \cdots, C_p$ are completely dependent on C_1.*

One can now introduce logical coordinates just as in the case of one domain of individuals and show again that Weyl's axioms are satisfied.

2) If there are several *independent disjoint domains* $C_1, \cdots, C_p$, i. e.,

---

[11] For logic we have to take axiom B (2) in its unlimited form, where a realisation of g may be any homomorphism of g onto a permutation group and *not* merely a matrix-representation.

no *a priori* functional dependence between elements of different domains then the calculus of propositional functions with arguments in one C_k only, has by Theorem 1.2 the symmetric group $\mathfrak{S}_{n_k}$ on C_k as its group of automorphisms. Propositional functions of r variables $x_j, x_k, \cdots$ from different domains $C_j, C_k, \cdots$ will transform according to the group induced in the set of all ordered r-tuples $(x_j, x_k, \cdots)$ by $\mathfrak{S}_{n_j}, \mathfrak{S}_{n_k}, \cdots$, i.e., according to the direct product $\mathfrak{S}_{n_j} \times \mathfrak{S}_{n_k} \times \cdots$. Invariance of the operations of the calculus of propositional functions (except disjunction and conjunction of propositional functions of variables from different domains) follows in the same way as before. To establish invariance of conjunction of propositional functions of arguments in different domains, let S_1, S_2 be arbitrary permutations of C_1, C_2 respectively and let x_1 be in C_1, x_2 in C_2. Then

$$f(S_1x_1) \,\&\, g(S_2x_2) = S_1f(x_1) \,\&\, S_2g(x_2) = S_1 \times S_2(f(x_1) \,\&\, g(x_2)),$$

where $S_1 \times S_2$ is the permutation induced by S_1 and S_2 in the set of all ordered pairs (x_1, x_2). Similarly for more arguments. Hence *the operations of the calculus of all those propositional functions, which are defined for every element of several a priori fixed disjoint domains $C_1, C_2, \cdots, C_p$ of individuals, are invariant under the symmetric groups on $C_1, C_2, \cdots, C_p$ and their direct products, provided the domains are completely independent* in the above sense.

In more complicated cases of several domains no such simple result will hold.

CHAPTER II.

Boolean Tensor Algebra.

The result obtained that logic is invariant-theory of the symmetric group reaffirms that there is an analogy of fundamental importance with geometry, especially with the linear geometries, and raises the question whether one cannot create a theory of invariants as an appropriate tool for the study of logical notions and their properties, similar to the powerful tools for the study of linear geometries, namely linear algebra (especially tensor algebra), theory of group-representations and invariants. This will now be attempted: First an appropriate "tensor-algebra."

1. Definition of a Boolean tensor. There is a great similarity between a propositional function of r arguments and a tensor of rank r. The propositional functions of one argument constitute the full Boolean algebra B_n of dimension n. Since a full Boolean algebra is uniquely determined (up to

isomorphism) by its dimension (= the cardinal of C), one can (just as in the case of vector algebra) take the set of all n-tuples of Boolean 0 and 1 as a model for B_n and define the operations of Boolean algebra to be performed by performing them on the "components of these Boolean vectors." This can be done for propositional functions by associating with every individual x in C in a one-one manner a "*dimension*" of a finite- or transfinite-dimensional Boolean vector space. Then the full Boolean algebra of all propositional functions $f(x_1, \cdots, x_r)$ can be isomorphically replaced by the Boolean algebra of all "*n-dimensional Boolean tensors of rank r*" if one *defines*: *An n-dimensional Boolean tensor $a_{j_1 \ldots j_r}$ of rank r is a single-valued function of r arguments ranging over all elements of a given set C of cardinal n, whose values are—relative to an arbitrarily fixed logical coordinate-system—Boolean 0 or 1 ("the components" of the Boolean tensor). Under a logical coordinate-transformation $a_{j_1 \ldots j_r}$ is transformed into*

$$a'_{k_1 \ldots k_r} = a_{Sj_1 \ldots Sj_r}$$

obtained by applying the permutation S to every index j. Alternatively,[12] if (s_{kj}) is the permutation-matrix corresponding to S

$$\begin{aligned} a'_{k_1 \ldots k_r} &= \sum_{j_1, j_2, \ldots, j_r} s_{k_1 j_1} s_{k_2 j_2} \cdots s_{k_r j_r} a_{j_1 j_2 \ldots j_r} \\ &= \prod_{j_1, j_2, \ldots, j_r} (\bar{s}_{k_1 j_1} + \bar{s}_{k_2 j_2} + \cdots + \bar{s}_{k_r j_r} + a_{j_1 j_2 \ldots j_r}). \end{aligned}$$

The last equality follows by observing that the permutation-matrix (s_{kj}) has in the row corresponding to the index i exactly one 1, at the intersection with the column i' say, and 0 elsewhere. Then $\sum_{k \in C} s_{ik} a_k = 1 \cdot a_{i'} = a_{i'}$. Hence $\prod_{k \in C} (\bar{s}_{ik} + a_k) = 0 + a_{i'}$, i. e. $\sum_{k \in C} s_{ik} a_k = \prod_{k \in C} (\bar{s}_{ik} + a_k)$ for any (finite or transfinite) permutation-matrix.

Thus the transformation law of a Boolean tensor is quite similar to that of an ordinary tensor; if a Boolean vector transforms according to the permutation S, i. e., according to the permutation matrix $[s_{jk}]$, then a Boolean tensor of rank r transforms according to the r-fold "Kronecker product" of the permutation matrix S:

$$[S]_r = [s_{j_1 k_1} s_{j_2 k_2} \cdots s_{j_r k_r}]$$

in the sense of the Kronecker product of matrices.

Since a permutation matrix is a special case of an orthogonal matrix, there

[12] We use $a + b, ab, \bar{a}$ for Boolean algebra-sum, product and complement respectively.

can be *no distinction between contravariant and covariant components in the case of Boolean tensors. Thus the Boolean principle of duality has—in contradistinction to geometry—its cause in a duality of the algebra of coordinates and not in a duality* (contragredience) *of the transformation group.* It also follows that there is no distinction between tensors and tensor densities here.

2. Addition, multiplication, subsumption and complementation of Boolean tensors. One has now to develop the possible invariant operations on Boolean tensors. They are quite analogous to those of ordinary tensor algebra, but owing to the fact that here the only transformations are permutations of the components, Boolean tensor algebra is richer in operations:

First one has the operations of Boolean algebra to be performed on the components of Boolean tensors of fixed rank r and yielding again Boolean tensors of rank r. E. g.,

$$a_{i_1 i_2 \ldots i_r} + b_{i_1 i_2 \ldots i_r} = c_{i_1 i_2 \ldots i_r}.$$

Besides, two Boolean tensors $a_{i_1 \ldots i_r}$, $b_{k_1 \ldots k_t}$ can be added (multiplied) by adding (multiplying) *every* component of a with *every* component of b, relative to a fixed logical coordinate-system:

$$a_{i_1 \ldots i_r} + b_{k_1 \ldots k_t} = f_{i_1 i_2 \ldots i_r k_1 k_2 \ldots k_t}, \quad a_{i_1 \ldots i_r} \cdot b_{k_1 \ldots k_t} = g_{i_1 i_2 \ldots i_r k_1 \ldots k_t}$$

yielding invariantly two Boolean tensors f, g of rank $r + t$, which we shall call the *outer sum* and *outer product of the Boolean tensors a and b*. One can, of course, form the outer sum (outer product) of any number of Boolean tensors. Outer sum (or product) of an infinite number of Boolean tensors would lead to Boolean tensors of infinite rank.

Example. Let a_i, b_k be two 3-dimensional Boolean vectors; their outer sum

$$[a_i + b_k] = \begin{pmatrix} a_1 + b_1 & a_1 + b_2 & a_1 + b_3 \\ a_2 + b_1 & a_2 + b_2 & a_2 + b_3 \\ a_3 + b_1 & a_3 + b_2 & a_3 + b_3 \end{pmatrix}$$

is a Boolean tensor of rank 2.

As a particular case of an outer product one can form the outer product of r Boolean vectors:

$$a_{i_1} \cdot b_{i_2} \cdot \ldots \cdot z_{i_r}$$

yielding a Boolean tensor of rank r. It follows from the transformation law of a Boolean tensor (II, 1) that *a Boolean tensor of rank r transforms like the outer product of r Boolean vectors,* just as for ordinary tensors.

However, not every Boolean tensor of rank r is of the form $a_{i_1} \cdot b_{i_2} \cdot \ldots \cdot z_{i_r}$, but it can always be expressed as sum of r-fold outer products of Boolean vectors (as is easily seen).

By duality, a Boolean tensor of rank r also transforms like the outer sum of r Boolean vectors:

$$a_{i_1} + b_{i_2} + \cdots + z_{i_r}$$

or as a combination of outer sums and products of r Boolean vectors e. g.:

$$a_{i_1} + b_{i_2} \cdot (c_{i_3} + \cdots + z_{i_r})$$

to which there is, of course, no analogy in ordinary tensor algebra.

Another invariant operation on Boolean tensors is the interchange of indices, e. g. f_{klmi} depends invariantly on f_{iklm}. In ordinary tensor space one derives invariant subspaces by imposing symmetry and anti-symmetry conditions. E. g.

$$a_{ikl} = a_{kil} \quad \text{or} \quad a_{ikl} = - a_{lki}.$$

In the case of a Boolean tensor one can still impose symmetry conditions, e. g. $f_{ijk} = f_{kji}$; however anti-symmetry conditions such as $f_{jik} = - f_{ijk}$ have no meaning here. Hence, whereas in ordinary tensor algebra there exists to every symmetry-class of tensors of rank r a complementary symmetry-class such that every (ordinary) tensor is the sum of two tensors, one from each class, this is not so for Boolean tensors.

The transformation law of a Boolean tensor $f_{i_1 i_2 \ldots i_r}$ completely symmetric in its indices (i. e., admitting every permutation of $i_1, i_2, \cdots, i_r$) is the same as that of the r-fold outer sum $f_{i_1} + f_{i_2} + \cdots + f_{i_r}$ or outer product $f_{i_1} \cdot f_{i_2} \cdot \ldots \cdot f_{i_r}$ of a Boolean vector with itself.

3. Contractions. Under the full linear group one can derive from an ordinary (mixed) tensor by "contraction" (i. e., by summation over two contragredient indices) a tensor of rank two less. Under the symmetric group one can already *sum over one single index*:

$$\sum_{j \in C} a_{i_1 \ldots i_{k-1} j i_{k+1} \ldots i_r}$$

obtaining invariantly a Boolean tensor of rank $r - 1$. Dually one can form

$$\prod_{j \in C} a_{i_1 \ldots i_{k-1} j i_{k+1} \ldots i_r}$$

Thus in Boolean tensor algebra there are *two kinds of contraction*: *Sum and product over one index, each lowering the rank of the Boolean tensor by one.*

Nevertheless one can contract over several indices simultaneously. E. g.:

$$\sum_{i \in C} a_{iij} = b_j \quad \text{or} \quad \prod_{i \in C} c_{ijiki} = d_{jk}$$

For, from a_{ikj} one obtains invariantly a Boolean tensor $f_{ij} = a_{iij}$ by putting $i = k$ in a and then contracting f_{ij}:

$$b_j = \sum_{i \in C} f_{ij} = \sum_{i \in C} a_{iij}$$

Besides putting indices equal one can *put indices unequal* in a Boolean tensor. E. g., let

$$b_{ik}(a_{ik}) = \begin{cases} a_{ik} & \text{for } i \neq k \\ \text{undefined} & \text{for } i = k. \end{cases}$$

Clearly, b_{ik} depends invariantly on a_{ik}. Similarly for Boolean tensors of higher rank.

Thus instead of contracting over all components of a_{ij} one can contract invariantly over all those for which $i \neq j$:

$$\sum_{i,j \in C} a_{ij} \ (i \neq j) \quad \text{or} \quad \prod_{i,j \in C} a_{ij} \ (i \neq j) \quad \text{or} \quad \sum_i \prod_j a_{ij} \ (i \neq j).$$

Or else over one index only. E. g.:

$$\sum_i a_{ij} \ (i \neq j), \qquad \prod_j a_{ij} \ (i \neq j).$$

4. Isomorphism with the calculus of propositional functions; Boolean tensor-equations. It is clear from the definitions of the various Boolean tensor-operations that *Boolean tensor algebra is isomorphic to the calculus of propositional functions over one* [13]—*a priori* fixed—*domain of individuals* such that to addition, multiplication, complementaton, subsumption and contractions of Boolean tensors there correspond conjunction, disjunction, negation, implication and quantification respectively of propositional functions. *Thus the calculus of propositional functions over one domain of individuals* [13] is seen to *be isomorphic to an invariantive calculus of greatest analogy to ordinary tensor algebra.*

Just as geometrical and physical laws can generally be expressed by tensor equations, so the properties of relations between the individuals of any mathematical theory which can be formalised within the calculus of propositional

[13] From I, 4 it follows that one can easily extend Boolean tensor algebra so as to be isomorphic to the calculus of propositional functions over several *fixed* domains C_k of individuals, if the transformations of each C_k are induced by the transformations of C_1.

functions over one[13] fixed domain of individuals could be expressed by invariant Boolean tensor equations.

In ordinary tensor algebra such invariant relations can be put into the form of the vanishing of a certain tensor expression, since the 0-tensor of rank r is left fixed under all linear transformations. In Boolean tensor algebra *many Boolean tensors of rank r are left fixed by $\mathfrak{S}_n$.* E.g. there are four fixed Boolean tensors of rank 2:

$$0_{ik} = 0 \text{ for all } i, k \text{ in } C;$$

$$\delta_{ik} = \begin{cases} 1 \text{ for } i = k \\ 0 \text{ for } i \neq k \end{cases} \text{(identity) and their duals } \bar{0}_{ik} = 1_{ik} \text{ and } \bar{\delta}_{ik}.$$

Similarly for Boolean tensors of higher rank. E.g.

$$f_{ijk} = \begin{cases} 0 \text{ for } i = j = k \\ 1 \text{ for } i = j \neq k \\ 0 \text{ for } i = k \neq j \\ 1 \text{ for } j = k \neq i \\ 1 \text{ for } i \neq j \neq k \neq i \end{cases}$$

is a Boolean tensor of rank 3 left fixed by $\mathfrak{S}_n$, and there are altogether 2^5 Boolean tensors of rank 3 left fixed by $\mathfrak{S}_n$, for any n. In fact it is easily seen that *all Boolean tensors of rank r left fixed by $\mathfrak{S}_n$ form a finite Boolean subalgebra of $[B_n]_r$* independent of the dimension n.

Any invariant Boolean tensor expression of rank r put equal to any one of these many fixed Boolean tensors of rank r will be of invariant meaning. However the most important among them, into which any other can be brought, are

$$g_{i_1 i_2 \ldots i_r} = 0_{i_1 i_2 \ldots i_r} \text{ or equivalently } \bar{g}_{i_1 i_2 \ldots i_r} = 1_{i_1 i_2 \ldots i_r}$$

where g stands for some invariant expression in Boolean tensors and

$$0_{i_1 i_2 \ldots i_r} = 0 \text{ and } 1_{i_1 i_2 \ldots i_r} = 1 \text{ for all } i_1, i_2, \cdots, i_r \text{ in } C$$

are the "*Boolean zero-tensor*" and the "*Boolean one-tensor of rank r*" respectively. And *any mathematical proposition which can be formalised within the calculus of propositional functions over one (or several completely[14] dependent) domain(s) of individuals can also be expressed as such an invariant equation.*

[14] In the sense of I, 4.

CHAPTER III.

Group-realisations.

Just as the theory of matrix-representations of (particularly linear) groups underlies the theory of invariants, tensor algebra and the development of a linear geometry as an invariant-theory, especially as it yields a knowledge of all linear transformation laws possible in the geometry in question, so one would have to study for the purpose of developing logic as invariant-theory the *realisations of a* (particularly the symmetric) *group by groups of arbitrary permutations.* But in the following we need not confine ourselves to groups:

DEFINITION. *A realisation R of an abstract semi-groupoid*[15] γ *is a homomorphism of* γ *onto a semi-groupoid* Γ *of permutations* S *of a set* C. The *degree* of R is the cardinal of C. The realisation R is called *faithful* if it is an isomorphism.

There is some analogy with the theory of matrix-representations: Since every matrix is an endomorphism of a vector space, one can reduce the theory of matrix-representations to the study of "representation-modules," i. e. vector spaces together with a domain of endomorphisms, i. e. groups with two domains of operators, and deduce from theorems on operator-groups theorems on matrix-representations, in particular Schur's lemma. Analogously, every permutation is by Theorem 1.1 an automorphism of a full Boolean algebra. Hence a realisation of a semigroupoid can be considered as a full Boolean algebra (or Boolean ring[16]) together with a domain of ring-endomorphisms (a *"realisation-module"*) and one can now deduce from theorems on (Boolean) rings with endomorphisms theorems on realisations, in particular an analogue of Schur's lemma (1).

And just as the regular (matrix-)representation of a finite group g is the source of all matrix-representations of g, so from the regular realisation (now considered as a permutation-realisation) all transitive realisations of g can be obtained.

[15] By a semi-groupoid γ is meant an algebraic system with a product $s_1 \circ s_2 =$ element of γ (not necessarily defined for all elements), which is associative whenever defined.

[16] Stone has shown that to every Boolean algebra there is a unique Boolean ring with identity and conversely; further that every homomorphism of a Boolean algebra is one of the corresponding Boolean ring and conversely (*Transactions of the American Mathematical Society,* vol. 40 (1935).

1. Realisation-modules. We now follow Emmy Noether's example and introduce a realisation-module in analogy to her representation-module:[17]

DEFINITION. Let $R: s \to S$ be a relisation of a semi-groupoid γ by means of permutations S of a set C and let B be the Boolean algebra (or ring) of all subsets of C. For every element s of $\mathfrak{S}$ and every b of B we define a product sb by means of the equation

$$sb = Sb.$$

We call this full Boolean algebra or ring with the second multiplication sb the *realisation-module corresponding to the realisation R.*

By Theorem 1.1 the mapping $b \to sb = Sb$ is an automorphism of the Boolean algebra (and ring) B, i. e. $s(b_1 + b_2) = sb_1 + sb_2$, $s(b_1 \cdot b_2) = sb_1 \cdot sb_2$, $s\bar{b} = \overline{sb}$.

In this way one can reduce the study of realisations of γ to the study of full Boolean algebras (or rings) together with a domain Ω of ring-endomorphisms. This (commutative) "Boolean Ω-ring" determines the realisation R completely by means of the equation $sb = Sb$ and conversely. For this purpose we need the notion of an *Ω-ring* which is analogous to (but not a special case of) an Ω-group. It is a special case of the general notion of an Ω-algebra, i. e. an algebraic system relative to a given domain Ω of its endomorphisms.

DEFINITION. *An Ω-ring is a ring $\mathfrak{R}$ together with a domain Ω of ring endomorphisms θ.* A subset of $\mathfrak{R}$ is called *admissible* or an *Ω-subset* if it is transformed into itself by every θ in Ω, which defines in particular *Ω-subrings* and *Ω-ideals*. Let $\mathfrak{R}$ and $\mathfrak{R}'$ be two Ω-rings with the same domain Ω. A mapping H of $\mathfrak{R}$ on $\mathfrak{R}'$ is called an *Ω-homomorphism* if it is a ring-homomorphism and if $H\theta = \theta H$ for every $\theta \in \Omega$.

To establish the following two theorems one has only to remark that they are true for Ω-groups and for ordinary rings.[18]

THEOREM 3.1 (*The Ω-homomorphism theorem for rings*). *If H is any Ω-ring-homomorphism of $\Omega - \mathfrak{R}$ on $\Omega - \mathfrak{R}'$ then $\mathfrak{R}'$ is Ω-isomorphic to $\mathfrak{R}/J$, where J is the Ω-ideal of all elements of $\mathfrak{R}$ mapped into $0'$, and conversely.*

COROLLARY. *The Ω-homomorphism theorem holds for Boolean rings and*

---

[17] Cf. v. d. Waerden, "Gruppen von linearen Transformationen," II, § 1 (*Ergebnisse der Mathematik* (1935)). Our treatment is strictly analogous to the one given there.

[18] I owe this observation to one of the referees.

Boolean algebras. For Stone has shown (*loc. cit.*) that the ordinary homomorphism theorem holds for Boolean rings and that every Boolean ring-homomorphism is a Boolean algebra-homomorphism and conversely.

THEOREM 3.2 (*The Ω-isomorphism theorem for rings*). *Let J_1 and J_2 be two Ω-ideals of an Ω-ring $\mathfrak{R}$, and let $J_1 + J_2$ be their greatest common divisor, $J_1 \cap J_2$ their least common multiple. Then*

$$(J_1 + J_2)/J_1 \cong J_2/(J_1 \cap J_2)$$

where $\cong$ means Ω-isomorphic.

In analogy to the application of the theory of Ω-groups to matrix-representations one can now apply Ω-rings to (permutation-)realisations:

1) Transitivity and intransitivity: If R is an intransitive (reducible) realisation, then C has a subset C_1 invariant under R. C_1 then transforms according to a realisation R_1 (say) called a *component* R_1 *of* R or *contained in* R. If B_1 is the Boolean subalgebra of B of all subsets of C_1 then $\Omega - B_1$ is a principal Ω-ideal of $\Omega - B$ and conversely every principal Ω-ideal of $\Omega - B$ defines a unique component of R. Hence the (distributive) lattices of all principal Ω-ideals of $\Omega - B$ and all components of R are isomorphic. In particular to the minimal principal Ω-ideals of $\Omega - B$ corresponds the irreducible components (subsets) of R (of C). Hence R is irreducible (= transitive if γ is a group) if and only if the corresponding realisation-module is simple, i.e. has no proper principal Ω-ideals.

2) If $\Omega - B$ is the *direct sum* of principal Ω-ideals $B_1, B_2, \cdots$

$$B = B_1 \dot{+} B_2 \dot{+} \cdots \text{ and therefore } R = R_1 \dot{+} R_2 \dot{+} \cdots$$

we say that $\Omega - B$ (and R) is *decomposable.* If C is finite or γ a group then the lattice of all principal Ω-ideals of $\Omega - B$ is a full Boolean algebra; hence in particular every group-realisation is sum of its transitive components (i.e. "*decomposes completely*"), as is well known.

3) Reduction into irreducible components: To successive reduction of a realisation R into irreducible components there corresponds by Theorems 3.1 and 3.2 the formation of a series

$$B = B_0 \supset B_1 \supset B_2 \supset \cdots \supset B_\nu = \{0\} \qquad (\nu = \text{any ordinal})$$

such that every B_λ contains its successor $B_{\lambda+1}$ as a proper maximal principal

Ω-ideal. If $\Omega - B$ decomposes completely the description can be made by a direct sum of minimal principal Ω-ideals.

4) Commutators: We first need

THEOREM 3. 3. *Every homomorphism of a finite Boolean algebra (or ring) B_n into a finite Boolean algebra B'_m can be expressed as a Boolean matrix M of m rows and n columns every row of which contains at most one* 1 *(and the rest* 0*'s).*

Proof. With respect to addition a Boolean ring of finite dimension constitutes an Abelian group with every element of order 2, all of whose homomorphisms are mappings $b_i \rightarrow \overset{n}{\underset{k=1}{\Delta}} a_{ik}b_k$ where Δ stands for Boolean ring-sum and b_k are the components of an n-dimensional vector over the field $F_2 = \{0, 1\}$, a_{ik} any matrix with m rows and n columns over F_2. In order to be a homomorphism also with respect to multiplication, i. e. $\Delta a_{ik}b_kc_k = (\Delta a_{ik}b_k)(\Delta a_{ik}c_k)$ it is necessary and sufficient that a_{ik} should have at most one 1 in each row. For suppose $a_{ik_1} = a_{ik_2} = 1$ and that $a_i = 0$ when j is neither k_1 nor k_2, then $\Delta(a_{ik}b_kc_k) = b_{k_1}c_{k_1}\Delta b_{k_2}c_{k_2}$ whereas

$$(\Delta a_{ik}b_k)(\Delta a_{ik}c_k) = (b_{k_1}\Delta b_{k_2})(c_{k_1}\Delta c_{k_2}) \neq b_{k_1}c_{k_1}\Delta b_{k_2}c_{k_2}$$

for all b and c unless either $a_{ik_1} = 0$ or $a_{ik_2} = 0$. But every Boolean ring homomorphism is also a Boolean algebra homomorphism (Stone, *loc. cit.*). Moreover since there is at most one 1 in each row $\Delta a_{ik}b_k = \sum\limits_k a_{ik}b_k$. I. e., *every homomorphism $b \rightarrow b'$ of B_n into B'_m is of the form $b \rightarrow b' = M; b$* where b, b' are the variable Boolean vectors in B_n, B'_m respectively and M is a Boolean matrix with at most one 1 in each row.

Hence if R_1 and R_2 are realisations of γ by permutations S_1, S_2 respectively and $\Omega - B_1$, $\Omega - B_2$ the corresponding realisation-modules, then any Ω-homomorphism can by Theorem 3. 3 be expressed by a Boolean matrix M with at most one 1 in each row such that $M; S_1(s) = S_2(s); M$ for all s in γ.

In particular if M is a permutation-matrix then $S_2(s) = M; S_1(s); M^{-1}$ i. e. equivalent realisations correspond to Ω-isomorphic realisation-modules and conversely.

If $R_1 = R_2$ then $M; S = S; M$, i. e. the Ω-endomorphisms of the realisation-module can be expressed by those Boolean matrices with at most one 1 in each row which commute with every S, and conversely.

One can now apply theorems on Ω-rings to realisations:

i) Uniqueness: If the finite chain condition holds for the lattice of principal Ω-ideals then the (generalised) Jordan-Hölder theorem entails uniqueness of successive decomposition into irreducible components. The Krull-Schmidt theorem can be applied to entail uniqueness (up to equivalence and order) of the directly indecomposable principal Ω-ideals of $\Omega - B$ and components of R.

ii) Analogue of Schur's lemma: Combining 4) above with Theorem 3.1 gives

THEOREM 3.4 (*Analogue of Schur's lemma*). *Let R_1 and R_2 be irreducible realisations of finite degree n, m respectively of a semi-groupoid γ:*

$$R_1\colon\ s \to S_1(s), \qquad R_2\colon\ s \to S_2(s).$$

Any Boolean matrix M of m rows and n columns with at most one 1 *in each row such that*

$$M\,;S_1(s) = S_2(s)\,;M \quad \textit{for all } s \textit{ in } \gamma$$

is either the zero-matrix or a permutation-matrix. In the latter case $m = n$ and R_1 and R_2 are equivalent.

In particular *every Boolean matrix M of this type which commutes with every member S of an irreducible family of permutations of a finite set is either the Boolean* 0*-matrix or a permutation-matrix.*

iii) Just as for a set of matrices one can now prove without difficulty

THEOREM 3.5. *Let Γ be a* (completely reducible) *system of permutations S of a finite set C: $\Gamma = \alpha_1\Gamma_1 + \alpha_2\Gamma_2 + \cdots + \alpha_r\Gamma_r$, meaning that the irreducible component Γ_k occurs α_k times in Γ (equivalent components being identified). Then*

$$M\,;S = S\,;M \quad \textit{for all } S \textit{ in } \Gamma$$

implies for any Boolean matrix M with at most one 1 *in each row that it must be of the form*

$$M = \begin{pmatrix} P_1 & 0 & 0 & \cdots & 0 \\ 0 & P_2 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & P_r \end{pmatrix}$$

where $P_1, P_2, \cdots, P_r$ are permutation-matrices or zero-matrices of degrees $\alpha_1 d_1, \alpha_2 d_2, \cdots, \alpha_r d_r$ respectively, d_k being the degree of the irreducible component Γ_k.

As a special case one can let M in Theorem 3.5 be a permutation-matrix and Γ a group. We then obtain as a particular case the well-known

COROLLARY. *A permutation M of a finite set C which commutes with every member of a group G of permutations of C must consist of cycles permuting only such elements of C as belong to equivalent transitive subsets of C with respect to G.* (By equivalent we mean such transitive subsets of C as transform according to equivalent transitive components of G.) Another special case is the well known fact that *the only permutations which commute with a given permutation S (both acting on the same finite set C) permute only such elements of C as belong to cycles of the same length of S.*

It should be noted that all the above argument about commutators (with at most one 1 in each row) of realisations only applies to realisations of finite degree, since every infinite Boolean algebra has non-principal ideals (Stone).

2. **Kronecker products; characters.** Just as the notion of the Kronecker product of matrices underlies that of an ordinary tensor, so the following notion of the Kronecker product of permutations underlies that of a Boolean tensor. Let S_1 be a permutation of a set C_1 of cardinal n_1, S_2 a permutation of a set C_2 of cardinal n_2 and $[S_{ik}^{(1)}]$, $[S_{ik}^{(2)}]$ the corresponding permutation-matrices, which can be considered to act on Boolean vectors a_i, b_j of dimensions n_1, n_2 respectively: $a_i \to \sum_k S_{ik}^{(1)} a_k$, $b_j \to \sum_l S_{jl}^{(2)} b_l$. Then the $n_1 \cdot n_2$ products $a_i b_j$ can be considered as the components of an $n_1 \cdot n_2$-dimensional Boolean vector (the outer product of a and b) which undergoes a transformation

$$a_i b_j \to \sum_{kl} S_{ik}^{(1)} S_{jl}^{(2)} a_k b_l$$

whose matrix $[S_{ik}^{(1)} S_{jl}^{(2)}]$ is the Kronecker product $[S_{ik}^{(1)}] \times [S_{jl}^{(2)}]$ of the permutation-matrices $[S_{ik}^{(1)}]$ and $[S_{jl}^{(2)}]$. Clearly $[S_{ik}^{(1)}] \times [S_{jl}^{(2)}]$ is again a permutation-matrix and the corresponding permutation $S_1 \times S_2$ is easily seen to be the permutation induced by S_1 and S_2 in the set $C_1 \times C_2$ of all ordered couples (x, y) with x in C_1, y in C_2. We call $S_1 \times S_2$ *the (Kronecker) product of the two permutations S_1 and S_2.* Then if S is the transformation of a Boolean vector, the corresponding transformation of a Boolean tensor of rank r is the r-fold product $S \times \cdots \times S = [S]_r$, say, of S with itself.

If one defines (as usual) the *character of a permutation* S of a set C to be the cardinal number of elements of C left fixed by S, then it follows at once that if $\chi_1(S)$, $\chi_2(S)$ are the characters of the two realisations R_1 and R_2

then $\chi_1(S)\cdot\chi_2(S)$ is the character of $R_1\times R_2$. It is well known that for a finite realisation of a finite group g of order $|g|$

$$\frac{1}{|g|}\sum_{s\in g}\chi(S)=\text{number of transitive components.}$$

But, by the above, the m-fold Kronecker-product $R\times\cdots\times R=[R]_m$ of R with itself is again a finite realisation of g, if R is, and it has the character $\chi^m(S)$. Hence *the character $\chi(S)$ of any realisation R of finite degree of a finite group g must satisfy the denumerable sequence of conditions*

$$\frac{1}{|g|}\sum_{s\in g}\chi^m(S)=\textit{positive integer},\qquad (m=1,2,3,\cdots).$$

In the same way it follows that *if $\chi_1(S),\chi_2(S),\cdots,\chi_\nu(S)$ are the characters of any finite number of realisations of finite degree of a finite group g of order $|g|$ then*

$$\frac{1}{|g|}\sum_{s\in g}\chi_1^{m_1}(S)\cdot\chi_2^{m_2}(S)\cdots\chi_\nu^{m_\nu}(S)=\textit{positive integer},$$

where $m_1,m_2,\cdots,m_\nu=$ any combination of positive integers.

3. The regular realisation. We first recall that every transitive realisation T of a group g is equivalent to a realisation obtained by associating with every group-element s of g the permutation which the left-cosets of a certain subgroup h of g experience under left-multiplication by s. We call h the subgroup which *generates* the transitive realisation T.[19]

We next need a criterion to decide whether a given transitive realisation T of a group g is contained in another given realisation D of g:

THEOREM 3.6. *Let $D: s\to S$ be a faithful realisation of a group g by means of a group G of permutations S of a set C and let T be a transitive realisation of g generated by the subgroup h of g. Suppose that under the realisation D there corresponds to h the subgroup H of G. Then T is contained in D if and only if there exists at least one element c of C such that H is the largest subgroup of G which leaves c fixed.*

Proof. Necessity: If T is contained in D, then there exists a subset C_1 of C which is transformed into itself according to T. T is generated by a certain subgroup h of g, i. e. there is a one-one correspondence

$$c_k\leftrightarrow w_k h$$

[19] Speiser, *Theorie der Gruppen*, § 37 (2nd edition).

between the elements c_k of C_1 and the left-cosets $w_k h$ of h in g, such that $Sc_j = c_k$ if and only if $sw_j h = w_k h$. The element c corresponding to h itself under this one-one correspondence is left fixed by every permutation S in H and by no other element of G, since $sh = h$ if and only if s is in h. Hence if T is contained in D then at least one element of the subset C_1 of C is left fixed by every element of the subgroup H and by no other element of G.

Sufficiency. Conversely, if to a given subgroup h of g there exists under the one-one correspondence $D: g \rightarrow G$ (and in particular $h \rightarrow H$) an element c_1 of C such that H is the maximal subgroup of G which leaves c_1 fixed, then any two permutations in G and not in H will transform c_1 into $c_2 \neq c_1$, if and only if they belong to the same left coset of H in G. Hence there is a 1-1 correspondence between the elements c_k of C into which c_1 is transformed and the left-cosets of H in G. Choose this 1-1 correspondence so that to $c_k = W_k c_1$ there corresponds the coset $W_k H$ of H in G. Then the element S of G transforms c_k into $Sc_k = SW_k c_1$ and $W_k H$ into $SW_k H$. Hence the transformations of the elements c_k of C (into which c_1 is transformed) constitute a transitive realisation which is equivalent to that generated by the subgroup H of G, which is in turn equivalent to that generated by the subgroup h of g, since the correspondence $g \rightarrow G$ is isomorphic.

THEOREM 3.7. *Let R be the regular realisation of any group g by means of permutations of a set C. Then the realisation $\langle R \rangle$ of g induced by R in the set $\langle C \rangle$ of all subsets of C contains all transitive realisations of g.*

Proof. We have only to show that the criterion of Theorem 3.6 is satisfied. Since R is the regular realisation we can identify C with g and the permutations of C with the left-translations of g. Then we have to show that there exists to every subgroup h of g at least one subset U of g, such that U goes over into itself under left-translations by elements of h, but by no other elements of g. Clearly any right-coset hw_k of h in g is such a subset $U: shw_k = hw_k$ if and only if s is in h.

If g is the symmetric group $\mathfrak{S}_n$ of all permutations of a finite set C, then the set of all ordered $(n-1)$-tuples $(x_1, x_2, \cdots, x_{n-1})$ of elements of C, such that $x_1 \neq x_2 \neq \cdots \neq x_{n-1}$, transforms according to the regular realisation of $\mathfrak{S}_n$. But it is a principal ideal of Boolean tensor-space of rank $n-1$. Hence we have

THEOREM 3.8. *Every transitive realisation of the finite symmetric group is contained in Boolean tensor space.*

By Theorem 3.5 this is equivalent to the fact that to every subgroup H

of $\mathfrak{S}_n$ there exists, if n is finite, at least one propositional function f, such that H is the largest subgroup of $\mathfrak{S}_n$ which leaves f fixed. As a natural generalisation of an automorphism of an algebraic system, we call *the maximal group of permutations of a set C which leaves a propositional function f, defined over C, fixed, the group of automorphisms of the propositional function f* [20] and have the result:

THEOREM 3. 9. *Every group of permutations of a finite set C is the group of automorphisms of at least one propositional function defined over C.*

However, the correspondence between propositional functions and their groups of automorphisms is many—one not one-one, as is easily seen.

If one admits propositional functions of transfinitely many variables (Boolean tensors of transfinite rank) then Theorems 3. 8 and 3. 9 would hold also for a domain with transfinitely many individuals.

CHAPTER IV.

Boolean Theory of Invariants.

1. Boolean quantities. In order to develop an analogue of the theory of invariants as the appropriate instrument for the study of logical notions and their properties our first step must be the *definition of "quantities"* which can appear as arguments and values in the formation of invariant functions. I. e., we have to define for every possible transformation law its corresponding substratum. Any such quantity q is reproducibly determined relative to a fixed logical coordinate-system by an m-tuple of Boolean 0's and 1's (an m-dimensional Boolean vector), the logical coordinates q_i (say) of q. In another logical coordinate-system q will have coordinates q'_i obtained from q_i by applying the permutation S to the components q_i which corresponds to the transition s from the first logical coordinate-system to the second. The transition s is an element of the given (abstract) group of automorphisms g, and S corresponds to s under the realisation $R: s \rightarrow S$; this is the transformation-law which characterises the "*kind of the quantity q*," whereas g characterises the theory as a whole. We can without loss of generality restrict ourselves first to those q which constitute full Boolean algebras. For any collection of q's, being determined relatively to any logical coordinate-system

[20] This definition has been given in the case of binary relations by O. Ore, Galois-connections, *Transactions of the American Mathematical Society* (1944).

by Boolean vectors, will constitute a subset of a full Boolean algebra. We are thus led to the following

DEFINITION. *A Boolean quantity q of kind R is an element of a full Boolean algebra B_m of a certain dimension m which is characterised by a realisation R of the group of automorphisms $g: s \to S$ of degree m. Each value of q determines relative to a logical coordinate-system an m-dimensional Boolean vector (q_i), which transforms under the transition s to another logical coordinate-system by applying the permutation S to the components q_i.*

Hence *a system of Boolean quantities of kind R is the realisation-module* (cf. III, 1) *corresponding to the realisation R of the group of automorphisms*; just as a system of quantities in geometry is a representation-module, i. e., a vector-space with a given matrix-representation induced in it, which describes the transitions to another (geometrical) coordinate-system. Thus "Boolean quantities" are strictly analogous to Weyl's [21] and v. d. Waerden's [22] quantities for geometry.

Although we need not specify the group of automorphisms in the definition of a Boolean quantity, for the purpose of logic it will be the symmetric group.

Examples. 1) The simplest case of a Boolean quantity is an atomic proposition, in extension. The realisation-module consists of the one-dimensional Boolean algebra $\{0, 1\}$ (the truth-values) with the identity-realisation as the induced group.

2) Boolean tensor-space of rank r is a system of Boolean quantities of kind $\mathfrak{S}_n \times \cdots \times \mathfrak{S}_n = [\mathfrak{S}_n]_r$ (= the r-fold Kronecker product of $\mathfrak{S}_n$ with itself).

3) Symmetrical Boolean tensors: All n-dimensional Boolean tensors of rank r whose indices satisfy certain symmetry-conditions form a full Boolean algebra with a certain realisation of $\mathfrak{S}_n$ induced in it.

Two systems of Boolean quantities will be called *equivalent* if they correspond to equivalent realisations.

A system Q of Boolean quantities q of kind R will in general contain other systems of Boolean quantities: Any full Boolean subalgebra of Q which is transformed into itself under R will be a system of Boolean quantities contained in Q. Among them are the admissible principal ideals of Q. To each

[21] *Classical Groups* (*loc. cit.*).

[22] *Mathematische Annalen*, vol. 113 (1937), p. 15.

minimal admissible principal ideal of Q there corresponds a transitive component of R. If Q has no proper principal admissible ideal we call it a *simple system of Boolean quantities.* I. e., *a simple system of Boolean quantities corresponds to a transitive realisation of g and conversely.*

Examples. 1) $f(x)$ is a simple quantity because $\mathfrak{S}_n$ is a transitive realisation of itself.

2) A Boolean tensor of rank $r > 1$ is not a simple Boolean quantity. E. g. a Boolean tensor f_{ik} of rank 2 contains two simple Boolean quantities, namely those f which are defined for $i = k$ and those for $i \neq k$ only.[23]

Defining the *direct sum* of Boolean quantities to be the direct sum of the realisation-modules which they constitute, it follows that Boolean quantities are "directly added" simply by juxtaposition of their components (just as for quantities with respect to linear groups). Hence *every system Q of Boolean quantities is the direct sum of simple subsystems of Q,* namely of those which correspond to the minimal admissible principal ideals of Q, *with respect to any group.*

Besides addition one can define the *Kronecker product of Boolean quantities*: If q_1 and q_2 are two Boolean quantities of kind R_1, R_2 respectively, then the Boolean quantity corresponding to $R_1 \times R_2$ will be called the *Kronecker product* $q_1 \times q_2$ of the two Boolean quantities q_1 and q_2. If $q_j^{(1)}$ and $q_k^{(2)}$ are the components of q_1 and q_2 then all possible sums $\Sigma q_j^{(1)} q_k^{(2)}$ range over all the components of $q_1 \times q_2$ and so do all the products $\Pi q_j^{(1)} + q_k^{(2)}$, as is easily seen.

Example. Boolean tensor space of rank r is the r-fold Kronecker product of Boolean vector space. Since the Kronecker product of two Boolean tensors is again one it follows that the class of simple systems of Boolean quantities occurring as admissible principal ideals of Boolean tensor space is closed with respect to Kronecker multiplication (followed by decomposition into admissible principal ideals). If Boolean tensor space of rank 1 belongs to this class, then it is obviously the smallest which is closed in this sense. However, not all simple Boolean quantities with respect to $\mathfrak{S}_n$ will belong to it in general. For examples show that the transitive components of $\mathfrak{S}_n \times \cdots \times \mathfrak{S}_n = [\mathfrak{S}_n]_r$ $(r = 1, 2, \cdots)$ do not exhaust the transitive realisations of $\mathfrak{S}_n$. Even in the simplest case of $\mathfrak{S}_3$ the realisation $\mathfrak{S}_3 \rightarrow \mathfrak{S}_2$ is not contained in $\mathfrak{S}_3 \times \mathfrak{S}_3$. To obtain all transitive realisations of $\mathfrak{S}_n$, if n is finite, one has to go further:

[23] Called self-relatives and alio-relatives by Peirce (*American Journal of Mathematics*, vol. 3, p. 44).

It is not sufficient to decompose Boolean tensor space into its principal admissible ideal, but to determine its transitive subsets. (That the latter is sufficient follows from Theorem 3. 8).

2. General properties of invariance. We are now confronted with the task of studying the possible invariant properties of and relations between systems of Boolean quantities. The first question must be: What is meant by an invariant property or relation?

DEFINITION OF INVARIANCE IN THE MOST GENERAL SENSE. *A property of or relation between members $x_1, x_2, \cdots, x_k$ of a finite number of sets $C_1, C_2, \cdots, C_k$ respectively, which can be expressed as a propositional function $f(x_1, x_2, \cdots, x_k)$, is invariant with respect to an abstract group g and realisations*

$$R_1 : s \rightarrow S_1(s),\ R_2 : s \rightarrow S_2(s),\ \cdots,\ R_k : s \rightarrow S_k(s)$$

of g as groups of permutations of $C_1, C_2, \cdots, C_k$ respectively if

$$f(x_1, x_2, \cdots, x_k) = f(S_1(s)x_1, S_2(s)x_2, \cdots, S_k(s)x_k)$$

for all s in g and every x_j in C_j $(j = 1, 2, \cdots, k)$.[24]

In the case of one set C only of cardinal n, all possible propositional functions $f(x)$ with x in C form the full Boolean algebra of dimension n. If, however, we limit our attention to those $f(x)$ which are invariant with respect to g and a realisation $R : s \rightarrow S(s)$ by permutations of C, then whenever $f(x)$ takes on a certain truth value for a certain x_0 in C it must have—by the above definition of invariance $f(x) = f(Sx)$—the same truth value for every x in C which belongs to the same set of transitivity as x_0. Also if $f(x_0) = f(y_0)$ and $g(x_0) = g(y_0)$ then $f(x_0) + g(x_0) = f(y_0) + g(y_0)$, $f(x_0) \cdot g(x_0) = f(y_0) \cdot g(y_0)$ and $\bar{f}(x_0) = \bar{g}(x_0)$ for any fixed x_0 and y_0 in C. Hence *the totality of invariant propositional functions $f(x)$ is isomorphic to the Boolean algebra of all subsets of the set whose elements are the transitive subsets of C with respect to g.* Similarly *the invariant propositional functions*

[24] In his *Grundlagen der Geometrie*, K. Reidemeister says: One could use the definition of a geometry as invariant theory of a group as a rigorous axiomatic definition if there existed a formal logical definition of invariance. Moreover such special notions as algebraic or differential in- and covariants are special cases of the above definition and so is the notion of an automorphism of an algebra.

This definition could at once be generalised to any (associative) semi-groupoid γ (instead of a group g) and realisations of γ as sets of arbitrary transformations of $C_1, C_2, \cdots, C_k$ (See the definition at the beginning of III). But no use will be made of this possibility in the following.

$f(x_1, x_2, \cdots, x_k)$ *constitute* a subalgebra of the Boolean algebra of all propositional functions of k arguments, namely *the full subalgebra consisting of those $f(x_1, x_2, \cdots, x_k)$ which take on the same truth value for any such ordered k-tuples $(x_1, x_2, \cdots, x_k)$ as are permuted amongst each other by g.* In short they constitute the subalgebra obtained by identifying any such ordered k-tuples as are permuted by g (under the realisations $R_1, R_2, \cdots, R_k$), which is possible since g is a group and therefore permutability under any realisation of g an equivalence-relation.

It follows that a knowledge of all invariant propositional functions $f(x_1, x_2, \cdots, x_k)$ is obtained by the decomposition of $R_1 \times R_2 \times \cdots \times R_k$ into its transitive components and conversely.

The invariant propositional functions $f(x_1, x_2)$ of two arguments are equivalently characterised by the condition that the Boolean matrix F (say) whose components are the values of $f(x_1, x_2)$ shall commute with the two realisations R_1 and R_2 in the sense that for every s in g we have

$$F; S_1(s) = S_2(s); F$$

where $S_1(s), S_2(s)$ are the permutation matrices corresponding to s under R_1, R_2 respectively. If $R_1 = R_2$ then the propositional functions $f(x_1, x_2)$ invariant under R are those Boolean matrices which commute with every S in R.

Returning to the case of k arguments, the number of invariant propositional functions can, if $g, C_1, C_2, \cdots, C_k$ are all finite, be obtained from the characters $\chi_1(s), \chi_2(s), \cdots, \chi_k(s)$ of $R_1, R_2, \cdots, R_k$ respectivley. For the character of $R_1 \times R_2 \times \cdots \times R_k$ is $\chi_1(s) \cdot \chi_2(s) \cdot \ldots \cdot \chi_k(s)$ and therefore the number of transitive components of $R_1 \times R_2 \times \cdots \times R_k$ is

$$m = \frac{1}{|g|} \sum_{s \epsilon g} \chi_1(s) \cdot \chi_2(s) \cdot \ldots \cdot \chi_k(s) \qquad (|g| = \text{order of } g)$$

and $2^m =$ number of invariant propositional functions $f(x_1, x_2, \cdots, x_k)$.

In particular the number of Boolean matrices which commute with two realisations R_1, R_2 (in the above sense) is

$$2 \exp\left\{ \frac{1}{|g|} \cdot \sum_{s \epsilon g} \chi_1(s) \cdot \chi_2(s) \right\}.$$

If $R = R_1 = R_2$ is a transitive realisation then the number of commuting Boolean matrices is

$$2 \exp\left\{ \frac{1}{|g|} \cdot \sum \chi^2(s) = 2^l \right\}$$

where $l =$ number of double-cosets of that subgroup in g which generates the realisation R (in the sense of III, 3). For it is known[25] that $\frac{1}{|g|} \sum \chi^2(s)$ is the number of double-cosets of that subgroup in g which generates the transitive realisation R. Since four Boolean matrices commute with $\mathfrak{S}_n$ and since $\mathfrak{S}_{n-1}$ is the subgroup of $\mathfrak{S}_n$ which generates the realisation of $\mathfrak{S}_n$ by itself (as follows at once from Theorem 3. 6) it follows that *the number of double-cosets of $\mathfrak{S}_{n-1}$ in $\mathfrak{S}_n$ is two, i. e.*

$$\mathfrak{S}_n = \mathfrak{S}_{n-1} + \mathfrak{S}_{n-1} a \mathfrak{S}_{n-1} (a \in \mathfrak{S}_n, a \notin \mathfrak{S}_{n-1})$$

if $n =$ finite > 2. Also *if the number of double cosets of any subgroup g' in any finite group g is 2 and the number of cosets of g' in g greater than 2, then the transitive realisation of g generated by g' has no self conjugate elements.* For if the number of double cosets is 2 then there are four commuting Boolean matrices, namely 0_{ik}, $\bar{\delta}_{ik}$, δ_{ik} and 1_{ik}, none of which except the identity δ_{ik} is a permutation-matrix, since their dimension is greater than 2.

Now suppose that g' is a subgroup of any group g and R a realisation of g by a group G of permutations of a set C. Under R there corresponds to g' a subgroup G' of G. Then the partition of C into transitive subsets with respect to G is a (not necessarily proper) refinement of the partition of C with respect to G. Hence we have

THEOREM 4. 1. *Given any set C of elements of invariant significance with respect to a group g, then C will also be a set of elements of invariant significance with respect to any subgroup g' of g and the propositional functions $f(x)$ [x in C] invariant with respect to g form a full Boolean subalgebra of the full Boolean algebra of the propositional functions $f(x)$ invariant with respect to g'.*

This gives an exact meaning to the statement: To decrease the group increases the possible invariants.

There is another way of "deriving" one invariant-theory from another which has been used much in geometry: If g' is a subgroup of g one considers instead of the g'-invariant propositional functions $f'(y)$ the g-invariant propositional functions $f(a, y)$ of one argument more, where the set $\{a\}$ contains an element a_0 such that $s \cdot a_0 = a_0$ if and only if $s \in g'$. (E. g. $g =$ projective group, $a_0 =$ plane at infinity, $g' =$ affine group). It is very easily seen that these two ways are equivalent for any two groups $g' \leqq g$ as far as arbitrary invariant propositional functions are concerned:

[25] Speiser, *Theorie der Gruppen,* § 39 (2nd edition).

THEOREM 4.2 (The "*adjunction argument*"). *Let $\mathfrak{A}$ and $\mathfrak{Y}$ be two realisations of a group g within sets $\{a\}$ and $\{y\}$. Suppose that g' is that maximal subgroup of g which leaves one element a_0 of $\{a\}$ fixed (under the realisation $\mathfrak{A}$). Then every g'-invariant propositional function $f'(y)$ is obtained from a g-invariant propositional function $f(a, y)$ by putting $a = a_0$, i. e. by putting $f(a_0, y) = f'(y)$.*

Proof. Consider all ordered pairs (a_0, y). Two pairs (a_0, y_1) and (a_0, y_2) will go over into each other under g if and only if there is an s in g' such that $s \cdot y_1 = y_2$.

3. Boolean invariants and covariants. Let us now apply the general definition of invariance of **2** to an invariant-theoretic study of Boolean quantities. I. e., we want to determine the possible invariant properties of, and relations between, Boolean quantities. Let $Q = \{\cdots, q, \cdots\}$ and $H = \{\cdots, h, \cdots\}$ be two systems of Boolean quantities of kind $R_1: s \rightarrow S_1(s)$ and $R_2: s \rightarrow S_2(s)$ respectively, with respect to a group g. A binary relation $f(q, h)$ between q and h will be invariant if $f(q, h) = f(S_1 q, S_2 h)$ for all s in g, h in H and q in Q, i. e. if $f: q \rightarrow h$ implies $S_1 q \rightarrow S_2 h$. (Using the same symbol f for the mapping $q \rightarrow h$, i. e. q is mapped into h if and only if $f(q, h) = 1$). Confining ourselves to single-valued mappings $f: q \rightarrow h$ we *define a Boolean quantity h of kind R_1 to be a "Boolean covariant" of another Boolean quantity q of kind R_2 if h is a single-valued function $h(q)$ (i. e. $f: q \rightarrow h$) such that*

$$fS_1(s) = S_2(s) f \text{ or equivalently } h(S_1 q) = S_2 h(q) \text{ for all } s \text{ in } g.$$

I. e., the mapping $q \rightarrow h$ shall commute with the two realisations R_1, R_2.

Similarly simultaneous Boolean covariants are defined.

As a special case, if H is one-dimensional and therefore R_2 the identity realisation then h ($= j$ say) is called a *Boolean invariant of q*. I. e., *a Boolean invariant $j(q)$ of q is a single-valued function $j(q)$ whose argument is a Boolean quantity q of kind R and whose value is Boolean 0 or 1 such that $j(q) = j(Sq)$ for all q and S.* Similarly a simultaneous Boolean invariant $j(q_1, q_2, \cdots)$ of several Boolean quantities is defined. Thus a Boolean invariant j is a special case of an invariant propositional function f in the general sense of **2**: The argument of j is a Boolean quantity, whereas that of f is arbitrary. *Hence the simultaneous Boolean invariants of Boolean quantities $q_1, q_2, \cdots$ are the invariant properties of and the relations between $q_1, q_2, \cdots$, while the Boolean covariants are those invariant relations which define a single-valued dependence.* I. e., the Boolean covariants $h(q)$ corre-

spond to a subclass of all simultaneous Boolean invariants $j(h, q)$, namely to those $j(h, q)$ which define a single-valued mapping $q \to h$.

Examples. 1) Let Q consist of all propositional functions $f(x)$ of one argument x. The two quantifiers $\sum_x f(x)$, $\prod_x f(x)$ are Boolean invariants of f. But so are any symmetric functions of the values of f.

2) A Boolean tensor f_{ik} of rank 2 has Boolean covariants $\sum_i f_{ik} = h_k$ and $\prod_k f_{ik} = h'_i$ for instance. But *quantifiers, though the simplest Boolean covariants*, are not the only ones: Clearly, any function of f_{ik}, symmetric in one index i, say, will be a Boolean vector $c_k(f)$ say, which is a Boolean covariant of f_{ik}. Any symmetric function in both indices i and k is a Boolean invariant of f. E. g. $\sum_{i_1 \cdots i_4} f_{i_1 i_2} f_{i_3 i_4}$.

THEOREM 4. 3. *Let h_1 and h_2 be two Boolean quantities of the same kind $R : s \to S$ and q be a Boolean quantity of kind $R' : s \to S'$. If h_1 and h_2 are Boolean covariants of q, then so are their Boolean sum, product and complement $h_1 + h_2$, $h_1 \cdot h_2$ and $\bar{h}_1$. Hence the Boolean covariants of fixed kind R of a Boolean quantity q form a Boolean algebra.*

Proof. $h_1(q)$ and $h_2(q)$ being Boolean covariants of q we have by definition

$$h_1(S'q) = Sh_1(q) \quad \text{and} \quad h_2(S'q) = Sh_2(q).$$

Then

$$h_1(S'q) + h_2(S'q) = Sh_1(q) + Sh_2(q).$$

But S is an automorphism of the system of Boolean quantities $h_1, h_2, \cdots$, therefore

$$Sh_1(q) + Sh_2(q) = S[h_1(q) + h_2(q)]$$

hence

$$S[h_1(q) + h_2(q)] = h_1(S'q) + h_2(S'q)$$

showing that $h_1(q) + h_2(q)$ is a Boolean covariant of q. Similarly it follows that

$$h_1(S'q) \cdot h_2(S'q) = [Sh_1(q)] \cdot [Sh_2(q)] = S[h_1(q) \cdot h_2(q)]$$

and

$$\bar{h}(S'q) = \overline{Sh}(q) = S\bar{h}(q)$$

i. e. that $h_1(q) \cdot h_2(q)$ and $\bar{h}(q)$ are Boolean covariants of q with h_1 and h_2.

Among the Boolean covariants of a system Q of Boolean quantities there

is a subclass of "*homomorphic Boolean covariants* $l(q)$" defined thus: If q_1 and q_2 are any two members of Q then

$$l(q_1+q_2)=l(q_1)+l(q_2),\ l(q_1\cdot q_2)=l(q_1)\cdot l(q_2) \text{ and therefore } l(\bar{q})=\bar{l}(q).$$

I. e., the mapping $q\rightarrow l$ is an Ω homomorphism of Q onto L (as defined in III, **1**). Hence by the corollary to Theorem 3.1 *a finite-dimensional simple Boolean quantity q has only such homomorphic Boolean covariants as are equivalent to q.* It also follows at once from the considerations of III, **1** that *all homomorphic Boolean covariants of a finite-dimensional Boolean quantity q are—up to equivalence—obtained by juxtaposition of the components of such simple Boolean quantities as constitute principal admissible ideals of Q.*

Example. f_{ii} is a homomorphic Boolean covariant of f_{ik} and so is f_{ik} $(i\neq k)$.

Since every minimal principal admissible ideal of Boolean tensor space consists of those Boolean tensors whose indices satisfy a maximum number of consistent equality and inequality conditions (as is easily seen), it follows that *all homomorphic Boolean covariants of a finite-dimensional Boolean tensor $f_{i_1i_2\ldots i_r}$ are—up to equivalence—obtained by imposing equality and inequality conditions on its indices.* In particular, *a finite-dimensional Boolean vector has only such homomorphic Boolean covariants as are equivalent to it.*

The *central problem* is now; *To obtain all Boolean in—and covariants of a given system of Boolean quantities.*

For Boolean invariants the general answer is almost trivial: *The Boolean invariants $j(q)$ of a system Q of Boolean quantities q constitute a full Boolean algebra isomorphic to the Boolean algebra of all subsets of the totality of transitive subsets of Q.* This follows at once from IV, **2**. In particular if g is of finite order $|g|$ and $\chi(s)$ the character of the realisation $\langle R\rangle$ induced in the set of all subsets of that set on which R acts ($R=$ kind of q), then the number of transitive subsets of Q is $\frac{1}{|g|}\sum_s\chi(s)$ and therefore the number of Boolean invariants $j(q)$ is $2\exp\left\{\frac{1}{|g|}\sum_s\chi(s)\right\}$.

In every given case one will, however, want to go a step further: To determine, if possible, a set of "*basic Boolean invariants*" explicitly as Boolean polynomials of the components of q, such that all Boolean invariants of q are generated by them. To show that there are always "Boolean basic invariants," we need

THEOREM 4.4. *Suppose that an n-dimensional full Boolean algebra B_n*

($n =$ arbitrary cardinal) is given as the n-dimensional Boolean vector space. Then any single-valued mapping μ of any subset U of B_n onto the one-dimensional Boolean algebra B_1

$$\mu : x \to \mu x = 0 \text{ or } 1 \qquad (x \text{ in } U)$$

can be expressed as a Boolean polynomial $P(x_k)$ of the components x_k of the Boolean vector x. By a Boolean polynomial of the x_k is meant any expression in the x_k obtained by Boolean addition, multiplication and negation (including transfinite sums and products).

Proof. The mapping μ determines a partition of the subset U of $B_n : U = U_1 + U_0$ such that

$$\mu x = \begin{cases} 1 \text{ if } x \in U_1 \\ 0 \text{ if } x \in U_0. \end{cases}$$

Every Boolean vector a in B_n determines a partition $K = K_1(a) + K_0(a)$ of the arbitrary index-set $K = \{k\}$ of cardinal n such that

$$a_k = \begin{cases} 1 \text{ for } k \text{ in } K_1(a) \\ 0 \text{ for } k \text{ in } K_0(a). \end{cases}$$

Now for every a in U_1 we construct the Boolean polynomial

$$P_a(x_k) = \prod_{k \in K_1(a)} x_k \cdot \prod_{k \in K_0(a)} \bar{x}_k$$

of the components x_k of x. Clearly

$$P_a(x_k) = \begin{cases} 1 \text{ if } x = a \\ 0 \text{ if } x \neq a. \end{cases}$$

We now take the sum

$$\sum_{a \in U_1} P_a(x_k) = P(x_k), \text{ say}$$

obtaining a Boolean polynomial $P(x_k)$ of the components x_k of the variable Boolean vector x in B_n with value 1 if x is in U_1 and 0 otherwise. I. e.

$$P(x_k) = \begin{cases} 1 \text{ if } \mu x = 1 \\ 0 \text{ if } \mu x = 0 \text{ or undefined.} \end{cases}$$

COROLLARY. *Every Boolean invariant $j(q)$ of an arbitrary Boolean quantity q with respect to any group g is a Boolean polynomial of the components of q.* Hence in particular *$j(q)$ is always a sum of certain Boolean polynomials of the components of q,* namely of those which have the value 1 for every element of one transitive subset of Q and 0 elsewhere.

We shall call these Boolean invariants of q of which all others are sums the *strict basic Boolean invariants of* q.[26] Their number is equal to the number of transitive subsets of Q, hence cannot always be finite. There may be smaller bases than this strict basis, such that every $j(q)$ is a Boolean polynomial of these basis invariants, but no longer a sum of them. However, even a smaller basis for the Boolean invariants of q will be finite if and only if the number of transitive subsets of Q is finite, because it is known that the number of Boolean polynomials generated by a finite number of symbols is always finite.[27] Hence *if the number of transitive subsets of Q is transfinite there can be no finite basis for the Boolean invariants of q.*

Any single-valued mapping $B_n \to B_m$ can, if both B_n and B_m are taken as Boolean vector spaces, be considered as m single-valued mappings $B_n \to B_1$ of the n-dimensional Boolean algebra B_n onto the one-dimensional, namely as the mappings of B_n onto each of the m components of the variable Boolean vector of B_m (m and n arbitrary cardinals). By Theorem 4.4 each of these m mappings can be expressed by a Boolean polynomial. Hence

THEOREM 4.5. *Any single-valued mapping $B_n \to B_m$ of an n-dimensional Boolean vector space into an m-dimensional Boolean vector space (m, n arbitrary cardinals) can be expressed by letting each component of the variable Boolean vector of B_m be a Boolean polynomial of the components of the variable Boolean vector of B_n. Hence, in particular, the components of any Boolean covariant of a Boolean quantity q are Boolean polynomials of the components of q.*

As an immediate consequence of Theorem 4.2 one has

THEOREM 4.6. *The adjunction argument holds for the Boolean invariants of any Boolean quantity q with respect to any two groups $g' \subseteq g$.*

Indeed one has only to repeat the remark made in IV, **3** that the Boolean invariants of q are the invariant propositional functions of q.

4. The Boolean invariants of propositional functions. A Boolean polynomial in the components x_k of a Boolean vector will be called *homogeneous of degree d* if it is the sum of monomials

$$\prod_{k \in K_d} x_k \qquad (K_d = \text{arbitrary index set of cardinal } d)$$

[26] By the principle of duality it follows at once that every $j(q)$ (except $j(q) = 1$ all q) can also be expressed as the product of certain Boolean invariants of q.

[27] *Lattice Theory*, Theorem 6.8 (p. 93).

each of degree d. It follows at once that every strict basic Boolean invariant of any Boolean quantity q of kind R is the product of a homogeneous Boolean invariant and the complement of a homogeneous Boolean invariant of q.

From now on we suppose that the *dimension of q is finite.* Then every homogeneous Boolean invariant of degree $\leq d =$ finite of q is of the form

$$\text{(i)} \qquad \sum_{k_1 \ldots k_d} P_{k_1 \ldots k_d} \cdot q_{k_1} \cdot \cdots \cdot q_{k_d}$$

where $P_{k_1 \ldots k_d}$ must be the components of a fixed member of the system of Boolean quantities of kind $R \times R \times \cdots \times R = [R]_d$. But the fixed members of $Q \times Q \times \cdots \times Q = [Q]_d$ form a full Boolean algebra each of whose atoms [28] is the Boolean sum of all members of one transitive subset of $[Q]_d$. Therefore every homogeneous Boolean invariant of degree $\leq d$ is the sum of certain expressions of the form (i) for each of which $P_{k_1 \ldots k_d}$ has a minimal number of values equal to 1. Since by the above every Boolean invariant of q is a Boolean polynomial of homogeneous one's, *these minimal homogeneous Boolean invariants form a basis for all Boolean invariants of a finite-dimensional Boolean quantity q.*

If q is a Boolean tensor $a_{ikl\ldots}$ of rank r then every homogeneous Boolean invariant (under $\mathfrak{S}_n$) of degree d of a is of the form

$$\sum P_{i_1k_1l_1 \ldots i_dk_dl_d \ldots} \cdot a_{i_1k_1l_1 \ldots} \cdot \cdots \cdot a_{i_dk_dl_d \ldots}$$

where $P_{i_1k_1l_1 \ldots i_dk_dl_d \ldots}$ must be the components of a *fixed* Boolean tensor of rank rd. But for such P to form the product $P_{i_1k_1l_1 \ldots i_dk_dl_d \ldots} \cdot \cdots \cdot a_{i_1k_1l_1 \ldots}$ is the same as to impose equality and inequality conditions on the indices of $a_{i_1k_1l_1 \ldots} \cdot a_{i_2k_2l_2 \ldots} \cdot \cdots \cdot a_{i_dk_dl_d \ldots}$; similarly for several Boolean tensors. *Hence every homogeneous Boolean invariant of finite-dimensional Boolean tensors $a_{ikl} \cdots$, $b_{ikl} \cdots$, $\cdots$ under $\mathfrak{S}_n$ is obtained by forming outer products, imposing equality and inequality conditions on the indices and summing over all indices (i. e. contracting).* By negation and multiplication all strict basic invariants are obtained from homogeneous ones, whence all invariants by addition. *Thus all Boolean Invariants under $\mathfrak{S}_n$ of Boolean Tensors of Finite Dimension (Propositional Functions over one finite Domain of Individuals) can be obtained by the processes of Boolean Tensor Algebra (Calculus of Propositional Functions).*

Incidentally, if one is interested in the Boolean invariants of a Boolean

[28] Meaning they cover 0.

tensor $a_{ikl}\cdots$ under a subgroup g of $\mathfrak{S}_n$ one has by the adjunction argument (Theorems 4. 2 and 4. 6) only to find a Boolean tensor f_0 which is left fixed by g and no larger subgroup of $\mathfrak{S}_n$ and to determine the simultaneous Boolean invariants $j(a,f)$ under $\mathfrak{S}_n$ and finally put $f=f_0$. That f_0 always exists follows from Theorem 3. 8 or 3. 9.

Applying the above to Boolean vectors yields at once the result that *the polarised elementary symmetric functions*

$$\sum a_{i_1}\cdot a_{i_2}\cdot\ldots\cdot b_{k_1}\cdot b_{k_2}\cdot\ldots\cdot c_{l_1}\cdot c_{l_2}\cdot\ldots \quad (\text{any two indices} \neq)$$

form a basis for the simultaneous Boolean invariants of finite-dimensional Boolean vectors $a_k, b_k, \cdots, c_k$ under $\mathfrak{S}_n$. In particular a basis for the Boolean invariants of one finite-dimensional Boolean vector consists of the *elementary symmetric functions*

$$\sigma_m = \sum_{k_1<k_2<\cdots<k_m} a_{k_1}a_{k_2}\cdots a_{k_m} \quad (m=1,\cdots,n)$$

where $\sum$ stands (as everywhere) for Boolean algebra-sum.

It is now easily seen that the Boolean algebra generated by $\sigma_1,\sigma_2,\cdots,\sigma_n$ has the following $n+1$ atoms:

$$\begin{aligned}
\psi_0 &= \bar a_1\bar a_2\cdots\bar a_n\\
\psi_1 &= a_1\bar a_2\cdots\bar a_n+\bar a_1 a_2\bar a_3\cdots\bar a_n+\cdots+\bar a_1\bar a_2\cdots\bar a_{n-1}a_n\\
&\cdots\cdots\cdots\cdots\\
\psi_m &= a_1a_2\cdots a_m\bar a_{m-1}\cdots\bar a_n+\cdots+\bar a_1\bar a_2\cdots\bar a_{n-m}\,\bar a_{n-m-1}\cdots a_n\\
&\cdots\cdots\\
\psi_n &= a_1a_2\cdots a_n.
\end{aligned}$$

Hence these $n+1$ Boolean polynomials constitute a strict basis for all Boolean invariants of a finite-dimensional Boolean vector $(a_1,a_2,\cdots,a_n)$ under the symmetric group $\mathfrak{S}_n$ in the sense that every non-zero Boolean invariant of a is a sum of some of them. This follows also directly by simply remarking that ψ_m has the value Boolean 1 if and only if a_k has exactly m components equal to 1.

Interpreting a_k as truth-values and therefore a as a propositional function $f(x)$ of one argument x with finite range, then ψ_m can be interpreted as the proposition: "$f(x)$ is true for exactly m values of x." Similarly σ_m becomes the proposition: "$f(x)$ is true for at least m values of x." These are the so-called numerical conditions.[29] The two sets of basic Boolean in-

[29] Hilbert-Bernays, *Grundlagen der Mathematik*, vol. I, p. 169 ff. (called *Anzahlbedingungen* there).

variants of a_k can therefore be interpreted thus: *Any proposition about one propositional function $f(x)$ of one argument x over a finite domain of individuals, which is at all of invariant meaning, can be expressed by numerical conditions.*

Another basis for the Boolean invariants of one finite-dimensional Boolean vector is obtained by applying the fundamental theorem on symmetric functions (which is true for any commutative ring with identity) to Boolean rings: *The elementary symmetric functions*

$$\mathop{\Delta}_{k_1<k_2<\cdots<k_m} a_{k_1}\cdot a_{k_2}\cdot\ .\ .\ .\ \cdot a_{k_m}$$

formed by means of Boolean ring-sum $a \Delta b$ (and product) constitute a basis for the Bolean $\mathfrak{S}_n$-invariants of one finite-dimensional Boolean vector.

5. On the possibility of an invariant-theoretic classification of mathematical theories. The result that logic is invariant-theory of the symmetric group places logic in a similar position for all mathematical theories (which are based on two-valued logic) to that of projective geometry for the linear geometries: The projective group is the largest linear group and therefore every other linear group a subgroup of it; hence any linear geometry can be "derived" from projective geometry by diminishing the group or, what amounts to the same, by demanding the additional invariance of one or several tensors. Similarly the symmetric group is the largest transformation group, the group of automorphisms of any mathematical theory its subgroup. This raises the question: Can one obtain mathematical theories from logic as geometries are obtained from projective geometry? I. e., can one identify mathematical theories as invariant-theories of subgroups of the symmetric group or as invariant-theories of the symmetric group relative to a "distinguished" Boolean tensor?

However, to be invariant-theory of a given group cannot lead to the definition of any mathematical theory unless an algebra of coordinates is given as well. For from IV, **2** it follows that to demand invariance means merely to consider *all* propositional functions with the same truth-values for arguments permuted into each other, hence mere invariance cannot give rise to any process of decision whether a propositional function is true or false for given individuals. Indeed if f is an invariant propositional function, then so is $\bar{f}$. Invariance is only an upper limit.

The question as to a further extension of Klein's Erlanger program must therefore be put thus: Can one identify any mathematical theory in the usual sense as invariant-theory of a subgroup of the symmetric group with two-

element Boolean algebra as algebra of coordinates? The answer seems to be in the negative. A subgroup of the symmetric group is in general the group of automorphisms of many propositional functions $f(x, \cdots)$, $g(x, \cdots, \cdots$ [cf. III, **3**], i. e., the group of automorphisms of any mathematical theory in which $f(x, \cdots)$ and/or $g(x, \cdots)$ and/or $\cdots$, are primitive ideas. Hence with B_1 as algebra of coordinates invariance with respect to a subgroup of the symmetric group would in general be ambiguous. Moreover the group of automorphisms of many mathematical theories is so small that it cannot characterise the theories in question. [E. g. The theory of numbers has only the identity as automorphism and so has the theory of real numbers. The complex numbers have only one automorphism besides the identity.] The way towards a group- and invariant-theoretic classification of mathematical theories seems to lie in a different direction:

Instead of considering subgroups of $\mathfrak{S}_n$, we keep $\mathfrak{S}_n$ as the underlying group. Indeed a mathematical theory in its completely formal abstract shape is the theory of individuals from whose nature one has entirely abstracted. I. e., (if we assume only one domain of individuals) any two individuals are a priori equivalent, indistingushable, i. e. any permutation of the individuals must be admissible. Hence *any* mathematical theory when completely formalised, must be invariant under the symmetric group $\mathfrak{S}_n$ of all permutations of its domain of individuals, although its group of automorphisms need not be $\mathfrak{S}_n$.

Example. Theory of numbers: The only automorphism is easily seen to be the identity. However if one puts Peano's axioms in the form where no special individual symbols with a distinguished meaning (e. g. zero) or special symbols for distinguished relation (e. g. successor or less than) occur, then it follows at once that each axiom is invariant under permutations of the individuals. Such a system of axioms has been given by Hilbert-Bernays.[30] One of these axioms is

$$Ad(a, b, c) \;\&\; Sq(b, r) \;\&\; Sq(c, s) \rightarrow Ad(a, r, s)$$

where Ad and Sq are arbitrary propositional functions of three and two variables respectively (interpretable as addition and successor) and a, b, c, r, s arbitrary individual variables. Let $f_{\kappa\lambda\mu}$, $g_{\nu\rho}$ be two arbitrary Boolean tensors of rank 3 and 2 respectively. Then the above axiom can be put in the form

$$\overline{f_{\kappa\lambda\mu} \cdot g_{\lambda\nu} \cdot g_{\mu\rho}} + f_{\kappa\nu\rho} = 1_{\kappa\lambda\mu\nu\rho} \quad (= \text{the "Boolean one-tensor" of rank } 5)$$

[30] Axiomen-system Z^{**} (*Grundlagen der Mathematik*, vol. I, pp. 465-467).

or

$$\prod_{\kappa,\lambda,\mu,\nu,\rho} (\bar{f}_{\kappa\lambda\mu} + \bar{g}_{\lambda\nu} + \bar{g}_{\mu\rho} + f_{\kappa\nu\rho}) = 1.$$

In this form the left-hand side is at once seen to be a simultaneous Boolean invariant of f and g with respect to $\mathfrak{S}_n$.

As another example take the axiom for mathematical induction in the form where the special symbol for zero has been eliminated, from the same system of axioms:

$$(z)\overline{Sq}(z,b) \;\&\; A(b) \;\&\; (x)(y)[Sq(x,y) \;\&\; A(x) \rightarrow A(y)] \rightarrow A(a)$$

where A is a variable propositional function of one argument. This can be put into the form

$$\sum_{\kappa,\lambda} (f_{\kappa\lambda} + \bar{A}_\lambda) + \sum_{\mu,\nu} f_{\mu\nu} \cdot A_\mu \cdot \bar{A}_\nu + \prod_\rho A_\rho = 1$$

in which the left-hand side is a simultaneous Boolean invariant with respect $\mathfrak{S}_n$ of the Boolean tensor f and the variable Boolean vector A.

In general the axioms of any completely formal axiomatic theory will—by the very nature of such a theory—have to be invariant under any permutation of the individuals, if there is only *one* domain of individual variables. By applying the universal quantifier to any free individual variables, the axioms become atomic propositions i. e., Boolean invariants under $\mathfrak{S}_n$ of the Boolean tensors which corerspond to the primitive ideas of the theory and (possibly) of one or several variable Boolean tensors. *Thus a knowledge of all simultaneous Boolean invariants under $\mathfrak{S}_n$ of propositional functions $f, g, \cdots$ would yield a systematic knowledge of all possible completely formalised axiomatic mathematical theories in whose axioms $f, g, \cdots$ occur as primitive ideas or variable propositional functions.* E. g. the determination of all Boolean invariants under $\mathfrak{S}_n$ of one propositional function of three variables would yield a knowledge of all those theories of which the theory of groups, the theory of quasi-groups etc. are special cases. The knowledge of all simultaneous Boolean invariants of two propositional functions, each of three variables, would give an enumeration of all those theories of which the theory of rings, the theory of integral domains or of fields etc. are special cases.

DUBLIN INSTITUTE FOR ADVANCED STUDIES,
DUBLIN, EIRE.

THE ADIABATIC LINEAR OSCILLATOR.*

By Aurel Wintner.

1. In the linear differential equation $x'' + fx = 0$, let the coefficient $f = f(t)$ be defined and continuous on an half-line, $t_0 < t < \infty$. The coefficient function and the integration constants will be restricted to the real field. Unless stated or implied otherwise, only the non-trivial solutions $x(t)$ (that is, those which do not vanish identically) will be considered.

In a posthumous note edited by Goursat, Fatou [1] has arrived, from two different points of departure, at the following conclusion: If $l > 0$ and $L < \infty$, where

(1) $$l = \liminf_{t\to\infty} f(t) \quad \text{and} \quad L = \limsup_{t\to\infty} f(t),$$

then

(2) $$x(t) = O(1) \quad \text{as } t \to \infty$$

holds for every solution of

(3) $$x'' + f(t)x = 0.$$

The first of Fatou's starting points suggesting this theorem is nothing but an appeal to Sturm's comparison theorem; $x'' + (L + \epsilon)x = 0$, where $L + \epsilon = \text{Const.} > L$, being of higher, and $x'' + (l - \epsilon)x = 0$, where $l - \epsilon = \text{const.} > 0$, of lower, frequency than (3), as $t \to \infty$. Somewhat more details are sketched with regard to the second approach. The latter consists in a formal introduction of "polar coordinates," $x = r\cos\theta$, and of splitting (3) into a non-linear differential equation of the second order for $r = r(t)$ and into a subsequent quadrature determing $\theta = \theta(t)$. This formalism is the same as that in Newton's method for determining the radius vector $r = r(t)$ of a path described under a central (conservative) force in an (x, y)-plane. Correspondingly, there is little doubt that this second approach suggested itself to Fatou in connection with his last long paper [2], dealing with astronomical models which depend on a non-central (conservative) force.

Both of Fatou's starting points, the "Sturmian" and the "Newtonian," will be utilized in the present paper, the problem of which is due to the fact that the assertion of Fatou's posthumous note is false (that is, (2) need *not* hold if the limits (1) are positive and finite, respectively). In order to see

* Received December 20, 1945.

this, it is sufficient to choose $f(t) = p + q \cos t$, where p and q are constants satisfying $0 < q < p$, and then to apply Poincaré's results concerning the stable or unstable nature of the characteristic exponents of the resulting Mathieu equation,

$$x'' + (p + q \cos t)x = 0; \tag{3 bis}$$

cf. [7]. Quite a different counter-example has been given by Perron [4], who, without emphasizing it, has disproved *more* than what has been claimed in Fatou's note. In fact, whereas $0 < q < p$ and (1) imply that $0 < l < L$ in the case (3 bis) of (3), a glance at Perron's example shows that the two positive limits (1) coincide in his case, although (2) is still false.

In order to avoid an interruption of subsequent calculations, let here be derived a slight simplification of Perron's counter-example, in a way which will admit extensions to more sophisticated situations. Curiously enough, the natural way of deriving counter-examples of any *possible* type (cf. below) proves to be precisely Fatou's "Newtonian" approach, $x = r \cos \theta$, with $\theta = t$.

If r is independent of t, then

$$x = r \cos t \tag{4}$$

is a solution of the pure harmonic equation $x'' + x = 0$. In order to violate (2), let $x'' + x = 0$ be replaced by

$$x'' + (1 + \phi)x = 0, \tag{5}$$

where $\phi = \phi(t)$ is a continuous function satisfying

$$\phi(t) \to 0 \text{ as } t \to \infty. \tag{6}$$

Then $f(t) = 1 + \phi(t)$ in (3) is a continuous function for which the two limits (1) coincide. But for any function $r = r(t)$ (of class C'', or having just a second derivative),

$$x'' + x = r'' \cos t - 2r' \sin t$$

is an identity by virtue of (4) alone. Hence, (4) will be a solution of (5) if $\phi(t)$ is defined by

$$-\phi = (r'' - 2r' \tan t)/r. \tag{7}$$

And what is now required is that $r = r(t)$ be so chosen as to make (7) a function which remains continuous (even when $\tan t$ becomes infinite) and satisfies (6).

The simplest choice of an $r = r(t)$ which fulfills these conditions but for which the function (4) violates (2) is the choice

$$r = t + \tfrac{1}{2}\sin 2t, \tag{8}$$

(which differs from Perron's example only in that it curtails the formal differentiations). In fact, (8) shows that $r' = 2\cos^2 t$. Hence, both $r'\tan t$ and r'' are continuous and bounded. But (8) also shows that $r \to \infty$ as $t \to \infty$. It follows therefore from (7) that $\phi(t)$ is continuous and satisfies (6). Nevertheless, (4) does not satisfy (3), since $r(t) \to \infty$.

2. There are various senses in which the linear oscillator (3), with a (positive) frequency $f^{\frac{1}{2}}$ which is a given continuous function of time, can be thought of as representing an *adiabatic* distortion of the conservative linear oscillator $x'' + \omega^2 x = 0$, with a constant frequency $\omega > 0$. From the point of view of Hilbert's space, the definition of an adiabatic behavior is represented by the condition

$$\int^{\infty} (f(t) - \omega^2)^2 dt < \infty$$

(to be satisfied by *some* $\omega = \text{const.} > 0$). But all that is available in this regard is contained in Weyl's linear perturbation theory in the (L^2)-space (cf. [5]). And his spectral theory of (3) deals, in the main, with *solutions* $x(t)$ satisfying the (L^2)-condition

$$\int^{\infty} x^2(t)\,dt < \infty,$$

whether the preceding (L^2)-condition, that concerning the *coefficient* $f(t)$ of (3), be satisfied or not; cf. [6].

Two other definitions of "$f(t)$ is close to $\omega^2 = \text{const.}$", namely,

$$f(t) \to \omega^2 \text{ as } t \to \infty$$

and

$$\int^{\infty} | f(t) - \omega^2 |\, dt < \infty,$$

were constrasted in [8]. Clearly, none of the three conditions implies either of the other two, the three conditions being based on the metrics of the respective function spaces (L^2), (L^∞), (L).

In [8], the (L^∞)-problem was just mentioned, in connection with the Poincaré-Perron theory of (3) in the case

$$\lim_{t\to\infty} f(t) = c \neq 0, \tag{9}$$

(where, however, c *must not* be of the "stable" form ω^2; cf. [3], pp. 158-160), and was formally contrasted with an (L)-theory. If the unit of time is so chosen that ω becomes 1, the result of this (L)-theory can be formulated as follows:

Suppose that $\phi(t)$ is a continuous function satisfying

$$\int^{\infty} |\phi(t)|\, dt < \infty. \tag{10}$$

Then, corresponding to every solution, $x(t)$, of (5), *there exists a solution, $a\cos(t-\alpha)$, of the trivial approximation, $x'' + x = 0$, to* (5) *in such a way that*

$$x(t) - a\cos(t-\alpha) \to 0 \text{ as } t \to \infty. \tag{11}$$

In addition, the formal differentiation of this asymptotic relation is legitimate, that is,

$$x'(t) + a\sin(t-\alpha) \text{ as } t \to \infty. \tag{12}$$

Conversely, if any solution of the trivial approximation, that is, any pair of constants a, α, is given, then (5) *has a unique solution $x(t)$ satisfying* (11) *and* (12).

Actually, only the first two of these three assertions were proved in [8]. However, the third assertion, that concerning the converse, is a corollary of the first two. In fact, let $x = x_1(t)$ and $x = x_2(t)$ be two solutions of (5). By the first two assertions, there exist pairs of constants, say a_1, α_1 and a_2, α_2, satisfying (11) and (12) for $x = x_1$ and $x = x_2$ respectively. But, since (5) is self-adjoint, the (binary) Wronskian of x_1 and x_2 is independent of t. If C denotes its value, the constant C is 0 only if x_1 and x_2 are linearly dependent. On the other hand, if the asymptotic formulae (11), (12) are substituted into the Wronskian of x_1 and x_2, it is seen that $C = 0$ if and only if $a_1 a_2 = 0$. Hence, the two solutions are linearly independent if and only of $a_1 \neq 0$ *and* $a_2 \neq 0$. Accordingly, the converse assertion of the theorem is clear from the principle of superposition.

It is worth nothing that the theorem dualizes the spaces (L^1) and (L^∞), the assumption (10) being in the former, whereas the assertions (11), (12) concern the latter. But this is quite an accident. In fact, easy examples show that the situation fails for conjugate spaces (L^p), (L^q) if $q = p/(p-1)$

> 1 (in fact, even if $p = q$). In addition, the theorem becomes false if (L^∞) and (L) are interchanged.

The purpose of the following considerations is to collect a few general facts, positive and negative, which result if this interchange of (L^∞) and (L) is effected; that is, if the (L^∞)-condition is the *assumption* for the perturbation, $\phi(t)$, of *the coefficient function of* (5), rather than the *assertion* for the perturbations, (11) and (12), of *the solutions of* (5). In this sense, what will be developed is the (L^∞)-theory, whereas the theorem italicized above contains the (L)-theory, of the adiabatic variations of a linear oscillator.

3. Let a function $x(t)$, defined and having a continuous second derivative on an half-line $t_0 < t < \infty$, be called *quasi-harmonic* if its graph imitates the behavior of the real non-trivial solutions,

$$a \cos \omega t + b \sin \omega t, \qquad (a^2 + b^2 > 0),$$

of $x'' + \omega^2 x = 0$, where $\omega^2 =$ const., > 0, in the following sense: There exists a sequence $t_0 < u_1 < u_2 < \cdots$ such that the distance $u_{n+1} - u_n$ tends, as $n \to \infty$, to a finite, positive limit $(= \pi/\omega)$, and $x = x(t)$ is positive and concave or negative and convex (from below) according as t is on the open intervals (u_1, u_2), (u_3, u_4), (u_5, u_6), $\cdots$ or (u_2, u_3), (u_4, u_5), (u_6, u_7), $\cdots$, the points $t = u_1, u_2, u_3, \cdots$ being both the zeros and the points of inflexion of $x = x(t)$. The properties of being concave or convex should be meant in their strict sense, that is, so as to exclude rectilinear segments. In particular, $x(t)$ is nowhere constant, and so there exists on each of the intervals (u_n, u_{n+1}) a unique point, say $t = v_n$, at which the derivative of $x(t)$ vanishes. The absolute value of $x(v_n)$ then is the maximum of $|x(t)|$ on the "half-wave" (u_n, u_{n+1}). By the "n-th amplitude," a_n, of $x(t)$ will be meant this (local) absolute maximum.

In the above notations,

(13) $$a_n = (-1)^{n+1} x(v_n) > 0, \qquad x'(v_n) = 0,$$

since

(14) $$\operatorname{sgn} x(t) = (-1)^{n+1} \text{ if } u_n < t < u_{n+1},$$

and

(15) $$u_n < v_n < u_{n+1}.$$

Correspondingly, the absolute value of the slope $x'(t)$ has on the interval (v_n, v_{n+1}) a unique maximum, say b_n, and

(16) $$x(u_n) = 0, \qquad b_n = (-1)^{n+1} x'(u_n) > 0,$$

the zeros $t = u_n$ being precisely the points of inflexion. The (local) absolute maximum, b_n, of $x'(t)$ will be referred to as the "n-th co-amplitude" of $x(t)$.

(i) *Every non-trivial solution $x(t)$ of (5) is quasi-harmonic.*

In (i), and in (ii), (iii), $\cdots$ below, the coefficient function, $\phi(t)$, occurring in (5) is meant to be a real-valued, continuous function satisfying the (L^∞)-condition (6), and $x(t)$ is any real-valued solution (which does not vanish identically).

In view of (6), it can be assumed that $1 + \phi(t)$ is positive on the half-line $t_0 < t < \infty$. Then (5) shows that $x(t)$ and $x''(t)$ have opposite signs, with the understanding that $x(t) = 0$ if and only if $x''(t) = 0$. This means that the curve $x = x(t)$ is turning its concavities toward the t-axis, with the understanding that the points of inflexion and only these points are on the t-axis. And $x(t)$ does not vanish identically; hence, by a standard property of the solutions of any differential equation (3), the roots t of $x(t) = 0$ cannot cluster, and so the same holds for the roots of $x''(t) = 0$. Consequently, the curve $x = x(t)$ cannot contain a rectilinear segment. This proves that the function $x(t)$ has all properties of a quasi-harmonic function, except possibly those requiring the existence of a sequence of zeros $u_1, u_2, \cdots$ and of a positive limit for the distance $u_{n+1} - u_n$.

In order to ascertain for $x(t)$ these properties also, let $\epsilon > 0$ and $t_\epsilon < t < \infty$. Then, if t_ϵ is large enough,

$$1 - \epsilon < 1 + \phi(t) < 1 + \epsilon, \qquad (\epsilon < 1),$$

by (6). Hence, (5) is "surrounded," in the sense of Sturm's oscillation theory, by two linear differential equations (3) the solutions of which are of the form

$$a \cos (t(1 \pm \epsilon)^{\frac{1}{2}} - \alpha),$$

where a and α are integration constants. But (if $a \neq 0$) the distance between two consecutive zeros of either of these harmonics is a function of ϵ, namely $\pi/(1 \pm \epsilon)^{\frac{1}{2}}$, which tends to π as $\epsilon \rightarrow 0$. And this limit process can be effected by letting the end-point of the half-line $t_\epsilon < t < \infty$ tend to ∞. It follows therefore from the simplest of Sturm's comparison theorems, that $x(t)$ has a sequence of consecutive zeros, say $u_1, u_2, \cdots$, and that

$$(17) \qquad u_{n+1} - u_n \rightarrow \pi \text{ as } n \rightarrow \infty;$$

π being the half of the common wave-length of the non-trivial solutions of the approximation $x'' + x = 0$ to (6).

This completes the proof of (i).

4. Somewhat deeper lies the following fact:

(ii) *If a_n denotes the n-th amplitude of a non-trivial solution $x(t)$ of* (5), *then, as $n \to \infty$,*

$$a_{n+1}/a_n \to 1. \tag{18}$$

Moreover, if b_n denotes the n-th co-amplitude of $x(t)$, then

$$a_n/b_n \to 1 \tag{19}$$

(hence $b_{n+1}/b_n \to 1$).

If $1 + \phi(t)$ in (5) is replaced by $\omega^2 + \phi(t)$, where ω is a positive constant, it is clear from (13) and (16) that the resulting change in the unit of time amounts to the replacement of 1 by $1/\omega$ on the right of the limit relation (19). In contrast, the 1 on the right of (18) has nothing to do with the choice, $\omega = 1$, of the t-unit in (5).

A proof of (ii) follows by applying a well-known device of Liapunov, rediscovered by G. D. Birkhoff (for references, cf. [8], pp. 424-425). It is the same device of square-sums on which Perron's theory [3] of the formally unstable case of (9) is based. In the case of (5), the square-sum of Liapunov has a simple meaning, since it can be interpreted as the energy, say $h = h(t)$, of the "varied" linear oscillator, calculated under the assumption that the presence of the small disturbance, $\phi(t)$, is neglected in the Hamiltonian function, but not in the solution, $x(t)$, of (5); so that

$$h = \tfrac{1}{2}(x'^2 + x^2). \tag{20}$$

Differentiation of (20), where $x = x(t)$ is the given solution of (5), gives $h' = xx'' + x'^2$. Hence, if x'' is substituted from (5), it follows that $h' = -xx'\phi$. But $|xx'| \leqq h$, by (20). Consequently, $|h'| \leqq h\,|\phi|$. This differential inequality for $h = h(t)$ and $\phi = \phi(t)$ can be written in the form $|(\log h)'| \leqq |\phi|$, since $h \neq 0$. In fact, (20) is positive throughout, since the simultaneous vanishing of x and x' (for some t) leads to that solution $x(t)$ of (5) which vanishes identically, that is, to the solution excluded in (ii).

The last differential inequality is equivalent to the estimate

$$|\log h(\beta)/h(\alpha)| \leqq \int_\alpha^\beta |\phi(t)|\, dt, \tag{21}$$

in which α ($> t_0$) and β ($> \alpha$) are arbitrary. And (21) readily leads to (ii). In fact, if either $\alpha = v_n$, $\beta = v_{n+1}$ or $\alpha = u_n$, $\beta = v_n$, it is seen from (15),

(17) and (6) that the integral on the right of (21) tends to 0 as $n \to \infty$. It follows therefore from (21) that

$$\log h(v_{n+1})/h(v_n) \to 0 \text{ and } \log h(v_n)/h(u_n) \to 0$$

as $n \to \infty$. But it is seen from (13), (16) and (20) that $h(v_n) = a_n^2$ and $h(u_n) = b_n^2$. Hence, the last formula line proves both assertions, (18) and (19), of (ii).

5. Of the same type as (ii) is the following fact:

(iii) *If c_n denotes the (absolute) area of the n-th half-wave of a non-trivial solution $x(t)$ of (5), then, as $n \to \infty$,*

$$c_{n+1}/c_n \to 1. \tag{22}$$

Moreover,

$$\tfrac{1}{2}c_n/b_n \to 1. \tag{23}$$

First, from (5),

$$x'(\alpha) - x'(\beta) = \int_\alpha^\beta x(t)\,dt + \int_\alpha^\beta x(t)\phi(t)\,dt,$$

where α and β are arbitrary. Hence, if

$$\alpha \to \infty, \qquad 0 < \beta - \alpha = O(1) \text{ and } \operatorname{sgn} x(t) = \text{const. when } \alpha < t < \beta,$$

then, acording to (6),

$$|x'(\alpha) - x'(\beta)| = \left|\int_\alpha^\beta x(t)\,dt\right| + o\left(\int_\alpha^\beta |x(t)|\,dt\right).$$

But (17), (15) and (14) show that the conditions required of α and β for the truth of this o-relation are satisfied if $\alpha = u_n$, $\beta = u_{n+1}$ and $n \to \infty$. And, according to (16), the expression on the left of the last formula line then becomes identical with the sum $b_n + b_{n+1}$, whereas the first integral on the right is precisely the value c_n defined before (22). Finally, the second integral, that following the o-sign, is majorized by $\beta - \alpha$ times the maximum of $|x(t)|$ on the t-interval (α, β) and so, since $(\alpha, \beta) = (u_n, u_{n+1})$, by $u_{n+1} - u_n$ times the n-th amplitude, a_n. Accordingly,

$$b_n + b_{n+1} = c_n + (u_{n+1} - u_n)\, o(a_n)$$

as $n \to \infty$. Hence, (23) follows from (19), (18) and (17). Finally, (19) and (18) show that (22) is a corollary of (23).

The truth of what would correspond to (17) in the same way as $a_{n+1}/a_n \to 1$ corresponds to $b_{n+1}/b_n \to 1$, that is, the truth of $v_{n+1} - v_n \to \pi$ (and even the existence of $\lim (v_{n+1} - v_n)$ as $n \to \infty$) remains undecided. All that will be shown is that

$$(24) \qquad 0 < \liminf_{n\to\infty} (v_{n+1} - v_n) \leqq \limsup_{n\to\infty} (v_{n+1} - v_n) \leqq \pi.$$

Clearly, the content of the first of the inequalities (24) can be described as follows:

(iv) *As $n \to \infty$, the abscissae of the amplitudes of a non-trivial solution $x(t)$ of (5) cannot cluster at the ends of the respective half-waves.*

This can be interpreted as a limitation of the way in which the uniform continuity of $x(t)$ on a bounded interval $t_0 < t < t^0$ can deteriorate as $t^0 \to \infty$. Needless to say, (i) and (ii) do not imply (and, in view of Perron's counter-example, cannot imply) that $x(t)$ remains bounded during this limit process. All that is true is that, as $t \to \infty$,

$$(25) \qquad x(t) = O(1) \text{ if and only if } x'(t) = O(1).$$

In fact, since $a_1, a_2, \cdots$ and $b_1, b_2, \cdots$ represent the local maxima of $|x(t)|$ and $|x'(t)|$ respectively, (25) is a corollary of (19).

If $t \to \infty$, then $x''(t) = -x(t) + x(t)o(1)$, by (5) and (6). Since a_n is the maximum of $|x(t)|$ on the n-th half-wave, it follows that $x''(t) = -x(t) + a_n o(1)$ holds uniformly in t if $v_n \leqq t \leqq v_{n+1}$ and $n \to \infty$; cf. (15), (17), and (18). Consequently, by Taylor's formula,

$$x(v_{n+1}) - x(v_n) = x'(v_n)(v_{n+1} - v_n) + \tfrac{1}{2}x''(v_n)(v_{n+1} - v_n)^2 + o(a_n)o(v_{n+1} - v_n)^2.$$

Hence, from (13),

$$a_{n+1} + a_n = \tfrac{1}{2}a_n(v_{n+1} - v_n)^2 + o(a_n)o(v_{n+1} - v_n)^2,$$

since $x''(v_n) = -x(v_n) + o(|x(v_n)|)$, by (5) and (6). If this representation of $a_{n+1} + a_n$ is divided by a_n, it is seen that the assumption that the values $(v_{n+1} - v_n)^2$ can cluster at 0 contradicts the relation (18). This contradiction proves the first, while (15) and (17) imply the last, of the inequalities (24).

6. It turns out that, in a certain sense, (i) and (ii) are of a final nature. For instance, it is clear from the Perron type of example which is defined by

(4) and (8) that, although the amplitude must satisfy (18), not only $a_n \neq O(1)$ but even $a_n \to \infty$ is possible. In the same example, $x(t)$ is seen to be the sum of $t \cos t$ and of a periodic function; so that $x(t)$ is certainly not almost-periodic in Besicovitch's sense.

In contrast, if the (L^∞)-condition, (6), is replaced by the (L^1)-condition, (10), then the n-th amplitude of $x(t)$ must tend to a finite (and non-vanishing) limit, which implies, among other things, that (2) cannot be violated. And $x(t)$ must be almost-periodic even in Weyl's restricted sense. All of this, and more, is contained in the theorem italicized after (9). On the other hand, nothing like (i) or (ii) can be true if (6) is replaced by (10). In order to see this, it is sufficient to observe that (10) can in no sense restrict the behavior of a continuous (or, for that matter, regular-analytic) function $\phi(t)$ on a sequence of t-intervals (t_1^*, t_1^{**}), $(t_2^*, t_2^{**}), \cdots$ satisfying $t_n^{**} - t_n^* \to \infty$ as $n \to \infty$.

The example (4) defined by (8) is of a somewhat primitive nature. In particular, its amplitudes *tend* to infinity. More pathological possibilities compatible with (6) can be realized by applying the following rule of construction:

Let $g(t)$, where $0 \leqq t < \infty$, be a real-valued function possessing a continuous first derivative and satisfying

$$(26) \qquad g(t) \to 0 \text{ and } g'(t) \to 0 \text{ as } t \to \infty$$

(neither g nor g' need be monotone). Then, if $G(t)$ denotes the "Fourier primitive,"

$$(27) \qquad G(t) = \int_0^t g(s) \cos s \, ds,$$

of $g(t)$, the function

$$(28) \qquad x(t) = e^{G(t)} \cos t$$

is a solution of a differential equation (5) in which $\phi(t)$ is a real-valued continuous function satisfying (6).

This rule supplies examples of various type, since, in view of the parenthetical remark following (26), the assumptions imposed on $g(t)$ can hardly restrict the oscillations of (27) or of the amplitudes, a_n, of (28) (the sequence $a_1, a_2, \cdots$ is that of the relative maxima of the absolute value of (28)).

In order to verify the rule, let $r = r(t)$, where $0 \leqq t < \infty$, be a real-

valued function which is of class C'' and has the following properties: The assignment (7) defines a function $\phi(t)$ which (notwithstanding the $\tan t$) is *continuous*, and such as to satisfy the (L^∞)-condition, (6). As was seen in the introduction, the function (4) defined by any such $r(t)$ is a solution of the differential equation (5) which belongs to the coefficient function (7). Hence, it is sufficient to verify that the conditions required of $r(t)$ are fulfilled by $e^{G(t)}$, if $G(t)$ is the "cosine primitive" of the function $g(t)$ which, in turn, is of class C' and satisfies (26). In fact, the rule can then be verified, just by identifying (4) with (28), as follows:

First, $G' = g \cos t$, by (27). Hence, if $r = e^G$,

$$r' = rg \cos t.$$

This representation of r' implies that

$$r'' = r\,(g^2 \cos^2 t + g' \cos t - g \sin t).$$

But the last two relations reduce (7) to

$$-\phi = g^2 \cos^2 t + g' \cos t + g \sin t;$$

so that the $\tan t$ becomes eliminated. Thus, $g(t)$ being of class C', the function $\phi(t)$ defined by the last formula line is continuous. And it satisfies (6), since (26) is assumed.

In view of the "converse" assertion following (12), no non-trivial solution of (5) can satisfy

(29) $$x(t) \to 0 \text{ as } t \to \infty,$$

if the continuous coefficient function is subject to (10). But if (10) is replaced by (6), it now follows that, *notwithstanding* (i), (ii), (iii), *a non-trivial solution of* (5) *can satisfy* (29). In fact, the preceding verification remains unaltered if $\cos s$, $\cos t$ in (27), (28) are changed to $\sin s$, $\sin t$ respectively. Hence, in order to satisfy (6) and (29), where $x(t) \neq$ const., it is sufficient to ascertain that, for a function $g(t)$ of class C',

(29 bis) $$\int_0^t g(s) \sin s \, ds \to -\infty \text{ as } t \to \infty$$

is not prevented by (27) alone. But this is seen from the example $g(t) = -\sin t/t$.

Since (29) is equivalent to $a_n \to 0$ as $n \to \infty$, this example proves the existence of the second possibility in the following assertion:

(v) *Put*

$$\lambda = \liminf_{n\to\infty} a_n \text{ and } \mu = \limsup_{n\to\infty} a_n, \tag{30}$$

where a_n denotes the n-th amplitude of a real, non-trivial solution $x(t)$ of a differential equation (5) *in which $\phi(t)$ is a continuous function satisfying* (6). *Then each of the seven possibilities*

$$0 < \lambda = \mu < \infty, \qquad \lambda = 0 = \mu, \qquad \lambda = \infty = \mu, \qquad 0 \leqq \lambda < \mu \leqq \infty$$

can actually occur.

The first of these possibilities is realized, of course, by $x'' + x = 0$, where $\phi = 0$. The third possibility is exemplified by (4) and (8). Needless to say, the third case, too, can be subordinated to the above rule. Correspondingly, the four remaining possibilities, that is, the "mixed" cases $0 \leqq \lambda < \mu \leqq \infty$, can be obtained by alternating the choice of $\phi(t)$ on consecutive t-intervals (of appropriately constructed lengths) in such a way that sometimes the second, and sometimes the third (or first), of the three "pure" types preponderates.

In view of (v), there are at most two ways of improving on (i), (ii), one being the adjunction to (6) of a metric assumption, (10), the other the restriction of $\phi(t)$ by some qualitative condition (possibly some restriction of monotony). In the latter regard, it seems to be a general principle that, by virtue of (6) alone, there is a certain *balance* between the asymptotic behaviors of two linearly independent solutions of (5), as illustrated by the following fact:

If (5) *has a* (*non-trivial*) *solution the amplitudes of which tend to* 0, *then* (5) *has another* (*linearly independent*) *solution the amplitudes of which tend to* ∞.

In other words, the occurrence of the second of the possibilities established by (v) necessitates, for one and the same differential equation (5) satisfying (6), the occurrence of the third of the possibilities. In fact, let $x(t)$ and $x^*(t)$ be two linearly independent solutions, and let a_n^*, b_n^*, u_n^*, v_n^* be the constants which belong to $x^*(t)$ in the same way as the constants a_n, b_n, u_n, v_n occurring in (13), (14), (15), (16) belong to $x(t)$. According to (13), the Wronskian of $x(t)$ and $x^*(t)$ at $t = v_n$ is $\pm a_n$ times the derivative of $x^*(t)$ at $t = v_n$. But a Wronskian belonging to any differential equation (3) is independent of t, and its (constant) value is 0 if and only if $x(t)$ and $x^*(t)$ are linearly dependent. Consequently, if $a_n \to \infty$, that is, if $x(t)$ satisfies the assumption, (29), then the absolute value of the derivative of $x^*(t)$

at $t=v_n$ must *tend* to ∞ as $n\rightarrow\infty$. Since the co-amplitude, $b_n{}^*$, of $x^*(t)$ is the maximum of the absolute value of the derivative of $x^*(t)$ between $t=v_n{}^*$ and $t=v_{n+1}{}^*$, it now follows from (15) and (17) that $b_n{}^*\rightarrow\infty$. In view of (18) and (19), this implies the assertion, $a_n{}^*\rightarrow\infty$. A corollary can be formulated as follows:

If (5) *has a* (*non-trivial*) *solution satisfying* (29), *then* (5) *has a* (*linearly independent*) *solution violating* (3).

It is also seen that (29) is equivalent to the corresponding assumption for $x'(t)$, that is,

$$x(t)=o(1) \text{ if and only if } x'(t)=o(1).$$

This corresponds to (25).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

[1] P. Fatou, "Sur un critère de stabilité," *Comptes Rendus*, vol. 189 (1929), pp. 967-969.

[2] ———, "Sur le mouvement d'un point matériel dans un champ de gravitation fixe," *Acta Astronomica* (Cracow), ser. a, vol. 2 (1931), pp. 101-164.

[3] O. Perron, "Ueber Stabilität und asymptotisches Verhalten der Integrale von Differentialgleichungssystemen," *Mathematische Zeitschrift*, vol. 29 (1929), pp. 129-160.

[4] ———, "Ueber ein vermeintliches Stabilitätskriterium," *Göttinger Nachrichten*, 1930, pp. 28-29.

[5] H. Weyl, "Ueber beschränkte quadratische Formen, deren Differenz vollstetig ist," *Rendiconti del Circolo Matematico di Paledmo*, vol. 27 (1909), pp. 373-392.

[6] ———, "Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen," *Mathematische Annalen*, vol. 68 (1910), pp. 220-269.

[7] A. Wintner, "On a statement of Fatou," *Journal of the London Mathematical Society*, vol. 11 (1936), pp. 245-246; cf. also "The periodic analytic continuations of the circular orbits in the restricted problem of three bodies," *Proceedings of the National Academy of Sciences*, vol. 22 (1936), pp. 434-439, where further references are given.

[8] ———, "Small perturbations," *American Journal of Mathematics*, vol. 67 (1945), pp. 417-430.

ON COMPONENTS OF A FUNCTION AND ON FOURIER TRANSFORMS.*

By H. KOBER.

1. Introduction. Let $F(\zeta)$ be defined on a Jordan curve C in the complex plane. Various writers have treated the problem of representing $F(\zeta)$ in the form $F(\zeta) = F_1(\zeta) + F_2(\zeta)$; the $F_j(\zeta)$ $[j = 1, 2]$ are required to be the limit-functions of functions $F_j(z)$ $[z = x + iy;\ z \to \zeta,\ \zeta$ on $C]$ which are analytic in the interior or exterior of C, respectively.[1]

Now replace C by the real axis. It is known that any function $F(x) \in L_2(-\infty, \infty)$ is representable in the form $F(x) = F_1(x) + F_2(x)$ where $F_1(x)$ or $F_2(x)$ is the limit-function of an element $F_1(z)$ of $\mathfrak{H}_2$ or $F_2(z)$ of $\bar{\mathfrak{H}}_2$ $[z = x + iy;\ y \to 0]$, respectively;[2] the class $\mathfrak{H}_p$ has been introduced by E. Hille and J. D. Tamarkin.[3] A similar result holds for $F(x) \in L_p(-\infty, \infty)$ $[1 < p < \infty]$; also for $p = 1$ and $p = \infty$, though the result is here less simple. There are, however, a number of further cases of interest, first of all the most general case $F(x)(1 + x^2)^{-1} \in L_1(-\infty, \infty)$. The problems thus arising will be treated in the present paper in detail, starting with preliminary results concerning the case where C is the unit circle.

The resolution of a function, defined on a Jordan curve C or on the real axis, into its components is a useful tool for analysis, for it is adapted to the reduction of problems on Lebesgue-integrable functions to problems on analytic functions. In fact, it has been applied in the approximation by rational

* Received November 4, 1945.

[1] E. g., J. Plemlj, *Monatshefte für Mathematik und Physik*, vol. 19 (1908), pp. 205-210; his method of proof, however, is not strict. J. L. Walsh, *Comptes Rendus*, vol. 178 (1924), pp. 58-59, and "Interpolation and approximation by rational functions in the complex domain," *American Mathematical Society Colloquium Publications*, 1935. A. Ghika, *Comptes Rendus*, vol. 186 (1928), pp. 1808-1810, and vol. 202 (1936), pp. 278-280; his results, and some more general and more detailed ones, can be deduced from 2 of the present paper by means of a conformal transformation.

[2] E. C. Titchmarsh, *Introduction to the Theory of Fourier Integrals*, Oxford 1937, Theorem 98.

[3] *Fundamenta Mathematicae*, vol. 25 (1935), pp. 329-352. $\mathfrak{H}_p$ is the class of functions $F(z)$ which, for $y > 0$, are regular and satisfy the inequality

$$\int_{-\infty}^{\infty} |F(x + iy)|^p\, dx \leq M^p \quad [0 < p < \infty] \quad \text{or} \quad |F(z)| \leq M \quad [p = \infty],$$

respectively. A function $F(z)$ is said to belong to $\bar{\mathfrak{H}}_p$ when $F(-z)$ belongs to $\mathfrak{H}_p$.

functions [4] and in the solution of integral equations [4] and, implicitly, in the theory of Fourier series. In the present paper the theory is employed to establish necessary and sufficient conditions for a function to be represented as a Fourier or Fourier-Stieltjes transform, using the Widder [5] theory of Laplace transforms. The theory is further used to deal with the Stieltjes transformation (**13-14**) and with Hilbert transforms (**15**), and to generalize the equation $2\cos z = e^{iz} + e^{-iz}$ (**6-7**).

The following notations will be used.[6]

$$(1.1)\quad Hf = H[f;\phi] = \frac{1}{2\pi}\,\text{P.V.}\int_{-\pi}^{\pi} f(e^{i\theta})\cot\tfrac{1}{2}(\phi-\theta)\,d\theta \qquad [\text{P.V.} = \text{Principal Value}].$$

$$(1.2)\quad \mathfrak{H}F = \mathfrak{H}[F(t);x] = \frac{1}{\pi}\,\text{P.V.}\int_{-\infty}^{\infty}\frac{F(t)\,dt}{t-x},$$

$$\mathfrak{K}F = \mathfrak{K}[F(t);x] = \frac{1}{\pi}\,\text{P.V.}\int_{-\infty}^{\infty}F(t)\left(\frac{1}{t-x}-\frac{t}{t^2+1}\right)dt.$$

$$(1.3)\quad f_j(z) = \frac{(-1)^{j-1}}{2\pi i}\int_{|\zeta|=1}\frac{f(\zeta)\,d\zeta}{\zeta - z} \qquad \left[\begin{matrix}|z|<1\\ |z|>1\end{matrix}\ \text{for}\ \begin{matrix}j=1\\ j=2\end{matrix}\right]$$

where the integral is taken over the unit circle, and

$$(1.41)\qquad F_j(z) = \frac{(-1)^{j-1}}{2\pi i}\int_{-\infty}^{\infty}\frac{F(t)\,dt}{t-z},$$

$$(1.42)\qquad \Phi_j(z) = \frac{(-1)^{j-1}}{2\pi i}\int_{-\infty}^{\infty}F(t)\left(\frac{1}{t-z}-\frac{t}{t^2+1}\right)dt,$$

where $z = x+iy$, and $y>0$ for $j=1$, $y<0$ for $j=2$; $\mathfrak{H}F$ is the Hilbert operator, and evidently

$$\mathfrak{H}F = \mathfrak{K}F + C_0, \qquad C_0 = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{tF(t)}{t^2+1}\,dt$$

whenever the latter integral and either $\mathfrak{H}F$ or $\mathfrak{K}F$ exist. Similarly

$$F_j(z) = \Phi_j(z) - (-1)^j C_0.$$

Constants depending on λ only are denoted by the single symbol A_λ.

---

[4] J. L. Walsh, *loc. cit.*, and H. Kober, *Proceedings of the Edinburgh Mathematical Society*, (2) vol. 7 (1946), pp. 123-133. I. Vecoua, *C. R.* (*Doklady*) *de l'Académie des Sci. de l'URSS*, vol. 26 (1940), 134, pp. 327-330.

[5] D. V. Widder, *The Laplace Transform*, Princeton, 1941.

[6] For the operators Hf and $\mathfrak{H}F$, see M. Riesz, *Mathematische Zeitschrift*, vol. 27 (1928), pp. 218-244. For $\mathfrak{K}F$, see H. Kober, *Journal of the London Mathematical Society*, vol. 18 (1943), pp. 66-71; there in the first footnote, p. 69, the "(3.2)" is to be omitted, the correct result is given in Lemma 11 of the present paper.

2. On functions defined on the unit circle. Except for Lemma 2, the results of this section are either known or, in substance, equivalent to known theorems on Fourier series.

LEMMA 1. *Let* $1 < p < \infty$ *and* $f(e^{i\theta}) \epsilon L_p(-\pi, \pi)$. *Then*

$$(2.1) \quad \int_{-\pi}^{\pi} | f_j(re^{i\theta}) |^p d\theta \leq A_p{}^p \int_{-\pi}^{\pi} | f(e^{i\theta}) |^p d\theta \quad \left[\begin{matrix} 0 < r < 1 \\ r > 1 \end{matrix} \ \textit{for} \ \begin{matrix} j = 1 \\ j = 2 \end{matrix}\right].$$

LEMMA 2. *Let* $f(e^{i\theta}) \epsilon L_1(-\pi, \pi)$ *and let* $0 < \lambda < 1$, $\nu < 1 - \lambda$. *Then*

$$(2.2) \quad \left(\int_{-\pi}^{\pi} \frac{| f_j(re^{i\theta}) |^\lambda d\theta}{| 1 + re^{i\theta} |^\nu}\right)^{1/\lambda} \leq A_{\lambda,\nu} \int_{-\pi}^{\pi} | f(e^{i\theta}) | \, d\theta \left[\begin{matrix} 0 < r < 1 \\ r > 1 \end{matrix} \ \textit{for} \ \begin{matrix} j = 1 \\ j = 2 \end{matrix}\right].$$

Lemma 1 is deduced from the inequality

$$\int_{-\pi}^{\pi} d\phi \, \Big| \int_{-\pi}^{\pi} g(\theta) \frac{e^{i\theta} + re^{i\phi}}{e^{i\theta} - re^{i\phi}} d\theta \Big|^p \leq A_p{}^p \int_{-\pi}^{\pi} | g(\theta) |^p d\theta \qquad \left[\begin{matrix} 0 < r < 1 \\ 1 < p < \infty \end{matrix}\right];$$

due to M. Riesz, taking $g(\theta) = f(e^{i\theta})$ and observing that

$$\int_{-\pi}^{\pi} g(\theta) \frac{e^{i\theta} + re^{i\phi}}{e^{i\theta} - re^{i\phi}} d\theta = - \int_{-\pi}^{\pi} f(e^{i\theta}) d\theta + 4\pi f_1(re^{i\phi}).$$

Evidently (2.1) holds for any p $[1 < p < \infty]$ if $f(e^{i\theta})$ belongs to $L_\infty(-\pi, \pi)$.

For $\nu = 0$ and $j = 1$ $[r < 1]$, (2.2) is in substance due to J. E. Littlewood.[7] From this result, (2.2) follows for $\nu = 0$ and $j = 2$ $[r > 1]$ by the identity

$$f_2(z) = -\frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{i\theta}) d\theta + \frac{1}{2\pi i} \int_{|\zeta|=1} \frac{f(\zeta^{-1}) d\zeta}{\zeta - z^{-1}} \qquad [| z | > 1].$$

To prove the general case we fix a number $q > 1$ such that $1 - \lambda > q^{-1} > \nu$ and take $\nu > 0$, $s = q(q-1)^{-1}$. By Hölder's inequality, the left side of (2.2) is not greater than

$$\left\{ \left(\int_{-\pi}^{\pi} | f_j(re^{i\theta}) |^{\lambda s} d\theta\right)^{1/s} \left(\int_{-\pi}^{\pi} \frac{d\theta}{| 1 + re^{i\theta} |^{\nu q}}\right)^{1/q} \right\}^{1/\lambda}$$
$$\leq A_{q,\lambda} \left(\int_{-\pi}^{\pi} | f_j(re^{i\theta}) |^{\lambda s} d\theta\right)^{1/(\lambda s)},$$

since $\nu q < 1$. We have $0 < \lambda s < 1$; applying the Littlewood result, we complete the proof.

[7] *Journal of the London Mathematical Society*, vol. 1 (1926), pp. 229-231.

Let $f(e^{i\theta}) \in L_1(-\pi, \pi)$. *Then, for almost all* ϕ *in* $(-\pi, \pi)$,

$$(2.3) \quad f_j(z) \to \tfrac{1}{2} f(e^{i\phi}) - \frac{i}{2}(-1)^j \{H[f;\phi] + \frac{1}{2\pi i}\int_{-\pi}^{\pi} f(e^{i\theta})\,d\theta\} \begin{bmatrix} z = re^{i\phi} \\ r \to 1 \end{bmatrix}.$$

This result is known. From the lemmas and from (2.3), we deduce

THEOREM 1. *Let* $1 \leq p \leq \infty$ *and* $f(e^{i\theta}) \in L_p(-\pi, \pi)$. *Then* $f(e^{i\theta})$ *can be represented in the form* $f(e^{i\theta}) = f_1(e^{i\theta}) + f_2(e^{i\theta})$ *where* $f_j(e^{i\theta})$ $[j = 1, 2]$ *is the limit-function of* $f_j(z)$ *as* $z \to e^{i\theta}$. *When* $1 < p < \infty$ *then the functions* $f_1(z)$ *and* $f_2(z^{-1})$ *belong to the Riesz class* H_p.[8] If $p = 1$ *or* $p = \infty$, *then they belong to* H_q *for any positive* q *smaller than* p, *while they belong also to* H_p *if and only if* Hf *belongs to* $L_p(-\pi, \pi)$.

The last assertion is deduced by means of Fatou's theorem and of the Smirnoff theorem:[9] If $0 < p < P \leq \infty$ and $g(z) \in H_p$, and if $g(e^{i\theta})$, the limit-function of $g(z)$, belongs to $L_P(-\pi, \pi)$, then $g(z) \in H_P$.

3. On functions defined on the real axis.

THEOREM 2. *Let* $F(x)(1 + x^2)^{-1} \in L_1(-\infty, \infty)$. *Then* $F(x)$ *can be represented in the form* $F(x) = F_1(x) + F_2(x)$ *where* $F_j(x)$ $[j = 1, 2]$ *is the limit function of* $\Phi_j(z)$ [*see* 1.4]. *Again*

$$(3.1) \quad \left\{ \int_{-\infty}^{\infty} \frac{|\Phi_j(x+iy)|^{\lambda} dx}{\{x^2 + (1 + |y|)^2\}^{\mu}} \right\}^{1/\lambda} \leq A_{\lambda,\mu} \int_{-\infty}^{\infty} \frac{|F(t)|\,dt}{1 + t^2} \begin{bmatrix} y > 0 & \text{for } j = 1 \\ y < 0 & \phantom{\text{for }} j = 2 \end{bmatrix}$$

for $0 < \lambda < 1$, $2\mu > 1 + \lambda$, *uniformly with respect to* y.

Some better results can be proved for the following particular cases:

(A) $F(x)(1 + |x|)^{-1} \in L_1(-\infty, \infty)$. (B) $F(x) \in L_1(-\infty, \infty)$.

(C) $F(x) \in L_p(-\infty, \infty)$, $1 < p < \infty$. (D) $F(x) \in L_\infty(-\infty, \infty)$.

It will be shown that $F_j(x)$ is the limit function of $F_j(z)$ or $\Phi_j(z)$, respectively, where $y > 0$ for $j = 1$; $y < 0$ for $j = 2$;

---

[8] F. Riesz, *Mathematische Zeitschrift*, vol. 18 (1923), pp. 87-95. A function $f(z)$ belongs to H_p if, for $|z| < 1$, it is regular and satisfies the inequality

$$\int_{-\pi}^{\pi} |f(re^{i\theta})|^p d\theta \leq M^p \quad [\theta < r < 1]$$

when $0 < p < \infty$; $|f(z)| < M$ when $p = \infty$. For the connection between the classes $\mathfrak{H}_p$ and H_p, see Hille-Tamarkin, *loc. cit.*, and H. Kober, *Bulletin of the American Mathematical Society*, vol. 49 (1943), pp. 437-443, Lemma 2.

[9] E. g., A. Zygmund, *Trigonometric Series*, Warsaw, 1935, 7.56 (iv).

$$\text{(3A)}\quad \left(\int_{-\infty}^{\infty}\frac{|F_j(x+iy)|^\lambda dx}{\{x^2+(1+|y|)^2\}^\mu}\right)^{1/\lambda}\leq A_{\lambda,\mu}\int_{-\infty}^{\infty}\frac{|F(t)|\,dt}{1+|t|}\begin{bmatrix}0<\lambda<1,\\ \mu>\frac{1}{2}\end{bmatrix};$$

$$\text{(3B)}\quad \left(\int_{-\infty}^{\infty}\frac{|F_j(x+iy)|^\lambda dx}{\{x^2+(1+|y|)^2\}^\mu}\right)^{1/\lambda}\leq A_{\lambda,\mu}\int_{-\infty}^{\infty}|F(t)|dt\begin{bmatrix}0<\lambda<1,\\ 2\mu>1-\lambda\end{bmatrix};$$

$$\text{(3C)}\quad \int_{-\infty}^{\infty}|F_j(x+iy)|^p dx\leq A_p{}^p\int_{-\infty}^{\infty}|F(t)|^p dt \qquad [1<p<\infty];$$

$$\text{(3D)}\quad \int_{-\infty}^{\infty}\frac{|\Phi_j(x+iy)|^q dx}{\{x^2+(1+|y|)^2\}^\mu}<\infty \qquad \begin{bmatrix}1<q<\infty,\\ \mu>\frac{1}{2}\end{bmatrix},$$
$$\leq A_q{}^q\int_{-\infty}^{\infty}\frac{|F(t)|^q dt}{1+t^2} \qquad \begin{bmatrix}1<q<\infty,\\ \mu\geq 1\end{bmatrix};$$

and the following theorems will be deduced:

THEOREM 2′. *Let $F(t)\in L_p(-\infty,\infty)$. When $1<p<\infty$ then $F_1(z)\in\mathfrak{H}_p$, $F_2(z)\in\bar{\mathfrak{H}}_p$. When $p=1$ then $F_1(z)\in\mathfrak{H}_p$ if and only if $\mathfrak{H}F\in L_1(-\infty,\infty)$; when $p=\infty$ then $\Phi_1(z)$ $\mathfrak{H}_p$ if and only if $\mathfrak{R}F\in L_\infty(-\infty,\infty)$; the conditions $F_1(z)\in\mathfrak{H}_1$ and $F_2(-z)\in\mathfrak{H}_1$ are equivalent and so are the conditions $\Phi_1(z)\in\Phi_2(-z)\in\mathfrak{H}_\infty$.*

THEOREM 2″ (*Uniqueness theorem*). (a) *Whenever a function $F(x)$ is represented in the form $F(x)=\bar{F}_1(x)+\bar{F}_2(x)$ where $\bar{F}_j(z)$ is the limit-function of $\bar{F}_j(z)$ $[j=1,2]$ and $\bar{F}_1(z)\in\mathfrak{H}_p$, $\bar{F}_2(-z)\in\mathfrak{H}_q$, $1\leq p\leq\infty$, $1\leq q\leq\infty$, then this representation is unique, except for a constant when $p=q=\infty$; in the latter case the function $\bar{F}_j(z)$ is equal to $\Phi_j(z)$ [see 1.42] except for a constant.* (b) *When $0<p<1$ and $0<q<1$, or when it is merely required that $\bar{F}_1(z)$ and $F_2(z)$ shall be regular for $y>0$ or $y<0$, respectively, then the representation is not unique.*

The last statement is evident. For the function $\{(z-1)(z-2)\cdots(z-n)\}^{-1}$ belongs both to $\mathfrak{H}_\lambda$ and $\bar{\mathfrak{H}}_\lambda$ whenever $n^{-1}<\lambda<1$.

Remark[10] *to Theorem 2.* Let (i) $F(x)$ and (ii) $\mathfrak{R}F$ be of bounded variation over $(-\infty,\infty)$. Then $\Phi'_1(z)=d\Phi_1(z)/dz$ belongs to $\mathfrak{H}_1$, $\Phi'_2(z)$ to $\bar{\mathfrak{H}}_1$, and $F(x)$ and $\mathfrak{R}F$ are absolutely continuous in $(-\infty,\infty)$; so is $F_j(x)$, being the integral of $\Phi'_j(x)=\lim\Phi'_j(z)$ $[y\to 0;\ j=1,2]$.

Remark to Theorem 2′. Let (i) $F(x)\in L_1$ and (ii) $\int_{-\infty}^{\infty}F(x)\,dx=0$.

[10] The proof is rather complicated. The statement can be considered as a generalisation of the Hille-Tamarkin Theorem 3.2 (i and iii).

Then there are functions $F_n(x)$ $[n = 3, 4, 5, \cdots]$ such that both F_n and $\mathfrak{H}F_n$ belong to L_1 and that $\int_{-\infty}^{\infty} | F(x) - F_n(x) | \, dx \to 0$ as $n \to \infty$. The conditions are necessary.[11]

Further remark. The inequalities (3.1), (3A), (3B), (3D) do not hold whenever $2\mu \leq 1 + \lambda$, $2\mu \leq 1$, $2\mu \leq 1 - \lambda$ or $2\mu \leq 1$, respectively.

To show this, take $F(t) = t^2 e(t)$, $te(t)$, $e(t)$ or $e(t)$, where $e(t) = 1$ for $0 < t < 1$, $e(t) = 0$ otherwise.

4. Proof of Theorem 2. By the transformation

$$(4.1) \quad t = \tan \tfrac{1}{2}\theta, \ \ x = \tan \tfrac{1}{2}\phi, \ \ z = i(1 - w)(1 + w)^{-1}, \ \ F(t) = f(e^{i\theta}) \qquad \left[-\pi < {\theta \atop \phi} < \pi\right]$$

we have

$$(4.2) \qquad f(e^{i\theta}) \in L_1(-\pi, \pi); \qquad \mathfrak{H}[F(t); x] = - H[f(e^{i\theta}); \phi].$$

$$(4.3) \quad \Phi_j(z) = f_j(w) + \tfrac{1}{2}(-1)^j a_0; \quad a_0 = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{F(t)\, dt}{t^2 + 1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{i\theta})\, d\theta.$$

From (2.3), (4.2) and (4.3) we deduce that, for almost all x in $(-\infty, \infty)$,

$$(4.4) \qquad \Phi_j(z) \to \tfrac{1}{2} F(x) + (i/2)(-1)^j \mathfrak{H} F \ \text{ as } y \to 0 \qquad [(-1)^j y < 0].$$

Thus $2F_j(x) = F(x) + (-1)^j i \mathfrak{H} F$, and so $F(x) = F_1(x) + F_2(x)$.

By a known argument, based upon a result due R. M. Gabriel,[12] we can now deduce (3.1) from (2.2), taking $\mu = 1 - \frac{1}{2}\nu$. In a similar way we prove (3D); when $\mu \geq 1$ we use (2.1), while for $\frac{1}{2} < \mu < 1$ we use the inequality

$$\int_{-\pi}^{\pi} \frac{| f_1(re^{i\theta}) |^q}{| 1 + re^{i\theta} |^\nu} \, d\theta \leq \left(\int_{-\pi}^{\pi} | f_1(re^{i\theta}) |^{qs} d\theta \right)^{1/s} \left(\int_{-\pi}^{\pi} | 1 + re^{i\theta} |^{-\nu l} d\theta \right)^{1/l}$$
$$\leq A_{q,r,s} \left(\int_{-\pi}^{\pi} | f(e^{i\theta}) |^{qs} d\theta \right)^{1/s} < \infty,$$

where

$f(e^{i\theta}) \in L_\infty(-\pi, \pi)$; $0 < \nu = 2(1 - \mu) < 1$; $0 < 1 - l^{-1} = s^{-1} < 1 - \nu$; $q > 1$.

To deduce (3A) from (2.2), we employ the transformation

$$(4.5) \quad t, x, z \text{ as in } (4.1); \ \ F(t) = (1 + e^{i\theta}) f(e^{i\theta}); \ \ F_j(z) = (1 + w) f_j(w);$$

---

[11] H. Kober, *Bulletin of the American Mathematical Society*, vol. 48 (1942), pp. 421-426, Theorem 3.

[12] *Journal of the London Mathematical Society*, vol. 5 (1930), pp. 129-131. Cf. Hille-Tamarkin, proof of Lemmas 2.1 and 2.5, *loc. cit.*

$$\mathfrak{H}[F;x] = (1+e^{i\phi})(ia_0 - Hf); \quad a_0 = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\theta})\,d\theta; \; \nu + \lambda = 2 - 2\mu.$$

To prove (3B), we replace (4. 5) by

(4. 6) t, x, z as in (4. 1); $\quad F(t) = (1+e^{i\theta})^2 f(e^{i\theta})$;

$$F_j(z) = (1+w)^2 f_j(w) + a_{-1}(1+w);$$

$$\mathfrak{H}F = (1+e^{i\phi})^2(ia_0 - Hf) + 2ia_{-1}(1+e^{i\phi});$$

$$a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\theta})e^{-in\theta}d\theta \;\; [n = 0, -1]; \qquad \nu + 2\lambda = 2 - 2\mu.$$

Some further results can be proved, for instance

THEOREM 2‴. *Let $(1+t^2)^{-1}F(t)$ or $(1+|t|)^{-1}F(t)$ belong to L_1. Then $(i+z)^{-2}\Phi_1(z)$ or $(i+z)^{-1}F_1(z)$ belongs to $\mathfrak{H}_1$ if and only if $(1+x^2)^{-1}\mathfrak{R}F$ or $(1+|x|)^{-1}\mathfrak{H}F$, respectively, belongs to L_1, and the condition $(i+z)^{-2}\Phi_1(z) \in \mathfrak{H}_1$ or $(i+z)^{-1}F_1(z) \in \mathfrak{H}_1$ is equivalent to $(i+z)^{-2}\Phi_2(-z) \in \mathfrak{H}_1$ or $(i+z)^{-1}F_2(-z) \in \mathfrak{H}_1$, respectively.*

Both in the cases (3A) and (3B) etc. we have, for almost all x in $(-\infty, \infty)$,

(4. 7) $$F_j(z) \to F_j(x) = \tfrac{1}{2}F(x) + (i/2)(-1)^j\mathfrak{H}F \qquad [y \to 0].$$

5. The inequality (3C) is, in substance, due to M. Riesz.[13] Together with (4. 7), it yields the first part of Theorem 2′. For the proof of the second part, we need

LEMMA 3. *Let $p = 1$ or $p = \infty$ and $F(t) \in L_p$. Then $F_1(z) \in \mathfrak{H}_1$ or $\Phi_1(z) \in \mathfrak{H}_\infty$, respectively, if and only if $\mathfrak{H}F \in L_1$ or $\mathfrak{R}F \in L_\infty$.*

Take first $p = 1$. The necessity of the condition follows from (4. 7) by Fatou's theorem. Conversely, let both F and $\mathfrak{H}F$ belong to L_1. Then the constant $a_{-1} = (4\pi)^{-1}\int_{-\infty}^{\infty} F(t)\,dt$ vanishes. Hence, by (4. 6), both $f(e^{i\theta})$ and Hf belong to $L_1(-\pi, \pi)$. By the last assertion of Theorem 1, therefore, $f_1(w)$ belongs to the Riesz class H_1. By (4. 6), we have $f_1(w)(1+w)^2 = F_1(z)$; employing a known result, we have $F_1(z) \in \mathfrak{H}_1$. The case $p = \infty$ is treated in a similar way, using (4. 1)-(4. 4).

[13] M. Riesz, *loc. cit.* The first part ($p = 1$) of Lemma 3 is proved by the author in the paper cited in [11]; but that method fails for $p = \infty$.

The proof of Theorem 2″ is based on known properties[14] of the "proper Cauchy integral" of an element of $\mathfrak{H}_p$ (cf. Lemma 9, **15**).

COROLLARY. *If $F(x)$ is represented in the form $F(x) = F_1(x) + F_2(x)$ where $F_1(x)$ and $F_2(-x)$ are limit functions of elements of $\mathfrak{H}_\infty$, then both $F(x)$ and $\mathfrak{R}F$ belong to $\mathfrak{H}_\infty$.*

Finally we state the following result:

If (i) $(1+|x|^m)^{-1}F(x) \in L_\infty(-\infty, \infty)$ $[m = 0, or = 1, or = 2, \cdots]$ *and* (ii) $\mathfrak{R}[(t+i)^{-m}F(t); x] \in L_\infty(-\infty, \infty)$, *then* $F(x) = F_1(x) + F_2(x)$ *where $F_j(z)$ is analytic for $y > 0$ $[j = 1]$ or $y < 0$ $[j = 2]$, respectively, and $|F_j(z)| < A(1+|z|)^m$ $[(-1)^{j-1}y > 0]$. The conditions are necessary; the components of $F(x)$ are unique except for an arbitrary polynomial of degree not greater than m.*

6. Let $G^{(\rho)}$ be the set of integral functions which are of order not greater than ρ, let $G_\alpha{}^{(\rho)}$ $[0 < \alpha < \infty]$ be the set of integral functions $F(z)$ such that, given any $\epsilon > 0$, $|F(z)| < A_{F,\epsilon} \exp\{(\alpha+\epsilon)|z|^\rho\}$. We shall prove

LEMMA 4. *Let $F(z) \in G^{(\rho)}$ or $\in G_\alpha{}^{(\rho)}$, respectively, let $F(t) \in L_p$, $1 \leq p \leq \infty$. Then $F_j(z)$ belongs to $G^{(\rho)}$ or $G_\alpha{}^{(\rho)}$ $[1 \leq p < \infty]$ and so does $\Phi_j(z)$ $[p = \infty]$; $[j = 1, 2]$.*

Let $1 < p < \infty$ and $F(z) \in G_\alpha{}^{(\rho)}$, and let C_r be the semi-circle $w = re^{i\theta}$, $-\pi \leq \theta \leq 0$. Then, for $y > 0$, we have

$$2\pi i\, F_1(z) = \int_{C_r} \frac{F(w)\,dw}{w-z} + \left(\int_{-\infty}^{-r} + \int_r^\infty\right) \frac{F(t)\,dt}{t-z} = I_1 + I_2 + I_3.$$

Since r is arbitrary, $F_1(z)$ can be continued analytically over the entire plane and is, therefore, an integral function. We have to show that, given $\delta > 0$, $|F_1(z)| < \exp\{(\alpha+\delta)|z|^\rho\}$. Let $\epsilon > 0$ and $(\alpha+\epsilon)(1+\epsilon)^\rho = \alpha+\delta$; we have $|F(w)| < A_{F,\epsilon} \exp\{(\alpha+\epsilon)|w|^\rho\}$. Let z be fixed, $|z| \geq 1$, and let $r = (1+\epsilon)|z|$. Then

---

[14] For $1 \leq p < \infty$ see Hille-Tamarkin, Theorem 2.1 (ii). For $p = \infty$ we define the "proper Cauchy integral" of a function $g(z)$ of $\mathfrak{H}_\infty$ by

$$G(z) = (2\pi i)^{-1} \int_{-\infty}^{\infty} g(t)[1/(t-z) - t/(t^2+1)]dt.$$

Then we have $G(z) = g(z) - \frac{1}{2}g(i)$ or $= -\frac{1}{2}g(i)$ for $y > 0$ or $y < 0$, respectively.

$$|I_1| \leq \pi(1+\epsilon^{-1})A_{F,\epsilon}\exp\{(\alpha+\delta)|z|^{\rho}\},$$

$$|I_j| \leq \left(\int_{-\infty}^{\infty}|F(t)|^p dt\right)^{1/p}\left(\int_r^{\infty}\frac{dt}{(t-|z|)^{p'}}\right)^{1/p'} < A_{F,\epsilon}\epsilon^{-1/p} \quad \begin{bmatrix} j=2,3 \\ p'=p(p-1)^{-1}\end{bmatrix}.$$

Hence $F_1(z) \in G_a{}^{(\rho)}$. The proof of the remaining assertions is similar.

7. We need not enlarge by formulating the theorems for $G^{(\rho)}$ and $G_a{}^{(\rho)}$ resulting from the lemma and the preceding theorems. We can, however, deduce some more detailed results for the class $G_a{}^{(1)}$.

THEOREM 3. *Let $F(z) \in G_a{}^{(1)}$ and $F(t) \in L_p$. (a) When $1 \leq p < \infty$ then*

$$(7.1) \qquad F(z) = F_1(z) + F_2(z); \quad F_j(z) \in G_a{}^{(1)}; \quad F_1(z) \in \mathfrak{H}_q, \; F_2(-z) \in \mathfrak{H}_q$$

for any q such that $p \leq q \leq \infty$ when $p > 1$, and such that $p < q \leq \infty$ when $p = 1$.

(b) *When $F(t)\log(t) \in L_1$* and $\int_{-\infty}^{\infty}F(t)\,dt = 0$, *then (7.1) holds for $q = 1$.*

(c) *When $p = \infty$, then $F_1(z)$ is bounded for $y > 0$, $F_2(z)$ for $y < 0$, if and only if the function*

$$(7.2) \qquad \Psi(x) = \int_{-\infty}^{\infty}F(t)\left(\frac{t}{t^2+1} - \frac{t-x}{(t-x)^2+1}\right)dt$$

is bounded in $(-\infty, \infty)$. This condition is certainly satisfied if

$$(7.3) \qquad \int_0^{x} F(t)\,dt = Ax + O(1) \qquad [x \to \pm\infty].$$

The first assertion follows from Theorem 2′, using Lemma 4 and the Plancherel-Pólya theorem: Let $F(z) \in G_a{}^{(1)}$ and $F(t) \in L_p$; then, for $p \leq q \leq \infty$, $F(t)$ belongs to L_q.[15] The hypotheses of Theorem 3(b) imply that $\mathfrak{H}F \in L_1$.[16] Using Lemma 3 we deduce (b). Now we have

$$\Re F + \frac{1}{\pi}\Psi(x) = \frac{1}{\pi}\int_0^{\infty}\{F(x+t) - F(x-t)\}\left(\frac{1}{t} - \frac{t}{t^2+1}\right)dt.$$

Let A_F be the upper bound of $F(x)$ in $(-\infty, \infty)$. Then $|F'(x)| \leq \alpha A_F$,[15] and so

$$|\Re F + \frac{1}{\pi}\Psi(x)| \leq \frac{2}{\pi}\alpha A_F\int_0^{\infty}\left(\frac{1}{t} - \frac{t}{t^2+1}\right)t\,dt = \alpha A_F \quad [-\infty < x < \infty].$$

[15] *Commentarii Math. Helv.*, vol. 10 (1937/38), pp. 110-163, § 30.

[16] H. Kober, *Journal of the London Mathematical Society*, vol. 18 (1943), p. 69.

Consequently the boundedness of $\Re F$ in $(-\infty, \infty)$ is equivalent to that of $\Psi(x)$. This gives the first part of (3c). Since $F(x)$ is bounded in $(-\infty, \infty)$, $\Psi(x)$ can be put into the form

$$\Psi(x) = -\lim_{M\to\infty} \int_{-M}^{M} \frac{F(x+t) - F(t)}{t^2+1} t dt.$$

Integrating by parts and using the hypothesis (7.3), we have

$$| \Psi(x) | = | \int_{-\infty}^{\infty} dt \frac{t^2-1}{(t^2+1)^2} \times \{ \int_0^{x+t} - \int_0^x - \int_0^t \} F(u) du | \leq \bar{A}_F \int_{-\infty}^{\infty} \frac{| t^2 - 1 |}{(t^2+1)^2} dt.$$

Hence (7.3) implies that $\Psi(x)$ is bounded, which completes the proof.

It can be shown that Theorem 3(c) holds when (7.3) is replaced by

$$(7.31) \qquad F(x) - Ax = \sum_{j=1}^{n} a_j x^{r_j} (\log x)^{\beta_j} + O(1) \quad \text{or} \quad \sum_{j=1}^{m} b_j | x |^{\tau_j} (\log | x |)^{\sigma_j} + O(1)$$

for $x \to \infty$ or $x \to -\infty$, respectively, where $0 \leq r_j < 1$, $0 \leq \tau_j < 1$; β_j, σ_j real.

The function $z^{-2}(1 - 2\cos z + \cos 2z)$ is an example for the case $1 \leq p = q \leq \infty$, while $\cos z$ is an example for the case $p = q = \infty$, with $A = 0$. The function $F_0(z) = \int_0^z u^{-1} \sin u du$, however, belonging to $G_-^{(1)}$, is bounded in $(-\infty, \infty)$, while the functions

$$(7.4) \quad \Psi(x) = -\int_{-\infty}^{\infty} \frac{tdt}{t^2+1} \int_t^{t+x} \frac{\sin u}{u} du = \frac{\pi}{2} \log | x | + O(1),$$

$$\Re F_0 = \int_0^x \frac{\cos u - 1}{u} du + \text{constant}$$

are not bounded; hence $F_0(z)$ cannot be re represented in the form (7.1), with $q = \infty$. Finally we note that the function

$$F(z) = (\tfrac{1}{4} + z^2)\Gamma(\tfrac{1}{4} + iz/2)\pi^{-\frac{1}{4} - \frac{1}{2}iz}\zeta(\tfrac{1}{2} + iz) \quad [\zeta(s) = \sum_1^{\infty} n^{-s}]$$

is an example for the class $G^{(1)}$ such that $F_1(z) \in \mathfrak{H}_p$, $F_2(z) \in \bar{\mathfrak{H}}_p$ whenever $1 < p \leq \infty$.

8. In the following sections we shall treat the *Fourier-Stieltjes* transform

$$(8.1) \qquad F(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-itx} d\alpha(t),$$

and the *Fourier transform* of a function $f(t)$ of $L_1(-\infty, \infty)$

$$(8.2) \qquad F(x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(t) e^{-itx} dt$$

and the $\mathfrak{M}^q$ Fourier transform of a function $f(t)$ of $L_p(-\infty, \infty)$ $[1/p + 1/q = 1, 1 < p \leq 2]$

$$(8.3) \quad F(x) = (2\pi)^{-\frac{1}{2}} \mathfrak{M}^q \int_{-\infty}^{\infty} f(t) e^{-itx} dt = (2\pi)^{-\frac{1}{2}} \operatorname*{lim.\,i.\,mean}_{N\to\infty}^{\text{order } q} \int_{-N}^{N} f(t) e^{-itx} dt.$$

In (8.1), $\alpha(t)$ is required to be a function of bounded variation over $(-\infty, \infty)$ and is supposed to be normalized, i.e. $\alpha(t) = \frac{1}{2}\{\alpha(t+0) + \alpha(t-0)\}$, $\alpha(0) = 0$. The following conditions are certainly necessary for a function $F(x)$ to be representable as a Fourier-Stieltjes transform: (i) The function

$$(8.4) \qquad h(t) = (2\pi)^{\frac{1}{2}} \lim_{M\to\infty} \int_{-M}^{M} F(x) e^{itx} dx$$

which should, in fact, be equal to $\alpha(t+0) - \alpha(t-0)$, exists for all t in $(-\infty, \infty)$, vanishes except at a finite or enumerable set $\{t_n\}$, and $\Sigma \mid h(t_n) \mid < \infty$, (ii) $F(x)$ is bounded and uniformly continuous in $(-\infty, \infty)$, and (iii) so is $\mathfrak{R}F$. For, by the previous theory, (8.1) implies that

$$i(2\pi)^{\frac{1}{2}} \mathfrak{R}F = \left(\int_0^{\infty} - \int_{-\infty}^{0}\right) e^{-itx} d\alpha(t) + A; A = \int_0^{\infty} e^{-t} d\{\alpha(t) + \alpha(-t)\}.$$

We note that (ii) implies the existence of functions $f_n(x)$ $[n = 1, 2, \cdots]$ represented as Fourier-Stieltjes transforms, even as Fourier transforms, uniformly bounded with respect to x and n, and such that $f_n(x) \to F(x)$ $[n \to \infty]$ for all x.[17]

These conditions, however, are not sufficient. The function $F_0(x) = \int_0^x u^{-1} \sin u du$ (see **7**) satisfies (i) and (ii), but not (iii) (see 7.4). Consequently it is not a Fourier-Stieltjes transform; the same is, therefore, true for $G(x) = e^{ix} F_0(x)$. Now $G(x)$ satisfies (i) and (ii). Since $G(z) \in \mathfrak{H}_\infty$, and since this implies that $\mathfrak{R}G = iG(x) - iG(i)$ [see 4.4 and [14]], the function $G(x)$ satisfies also (iii).

[17] Compare this with a fundamental property of Fourier-Stieltjes transforms in the case when both $\alpha(t)$ and the $\alpha_n(t)$ are required to be real monotonous functions. E.g. S. Bochner, *Vorlesungen über Fouriersche Integrale,* Leipzig 1932, Theorem 21.

9. To solve the problem thus arising for *Fourier-Stieltjes transforms* and, correspondingly, for *Fourier transforms,* we have to use the Widder theory of Laplace transforms. Introducing the operator

$$(9.1)\quad g_k(t) = g_k[F(x)\,;t] = \int_{-\infty}^{\infty} \frac{F(x)\,dx}{(1-itx/k)^{k+1}} \qquad [k=1,2,3,\cdots],$$

we obtain the following results.

THEOREM 4. *The function $F(x)$ is, for almost all x in $(-\infty,\infty)$, equal to* (A) *a Fourier-Stieltjes transform, or* (B) *to the sum of a constant and of the Fourier transform of an element of $L_p(-\infty,\infty)$ $[1<p\leq 2]$, respectively, if and only if*

$$(9.2)\ (\mathrm{A})\ \int_{-\infty}^{\infty}|\,g_k(t)\,|\,dt < M \ \text{ or }\ (\mathrm{B})\ \int_{-\infty}^{\infty}|\,g_k(t)\,|^p dt < M^p\ [k=1,2,\cdots]$$

where M is independent of k.

THEOREM 4′. *The function $F(x)$ is, for almost all x in $(-\infty,\infty)$, equal to the sum of a constant and of the Fourier transform of an element of $L_p(-\infty,\infty)$ $[1\leq p\leq 2]$ if, and only if, $\{g_k(t)\}$ $[k=1,2,\cdots]$ is a weakly convergent sequence in $L_p(-\infty,\infty)$.*

Trivially the weak convergence of $\{g_k(t)\}$ in L_1 is a much stronger condition than (9.2) (A), and obviously there is strict equality in the cases concerned in Theorem 4(A) and Theorem 4′ $(p=1)$, if $F(x)$ is required to be continuous in $(-\infty,\infty)$. The constants are zero if we add the conditions $F(x)\to 0$ as $x\to\infty$ for $p=1$, $(1+x)^{-1}F(x)\in L_1(0,\infty)$ for $1<p\leq 2$.

For completeness we state results which have been proved in another paper.[18]

[18] H. Kober, *Journal of the London Mathematical Society,* vol. 19 (1944), pp. 144-152. In this paper Theorems 4 and 4′ have been stated without proof. The condition of weak convergence of $\{g_k(t)\}$ in L_1 $(-\infty,\infty)$ can be replaced by the following one: given $\epsilon>0$, there is a $\delta=\delta(\epsilon)>0$ such that, uniformly with respect to k $(k=1,2,\cdots)$. $|\int_E g_k(t)\,dt\,|<\epsilon$ whenever the set E satisfies the condition $|\int_E [dt/(1+t^2)]|<\delta$. There is an analogous result for Fourier series, better than the known ones with respect to sufficient conditions: Let $0<r<1$,

$$s_n(x)=\sum_{k=-n}^{n} a_k e^{ikx},\qquad \sigma_n(x)=[1/(n+1)]\sum_{j=0}^{n} s_j(x),\qquad u(r,x)=\sum_{-\infty}^{\infty} a_n r^{|n|}e^{inx}.$$

Then $\Sigma a_n e^{inx}$ is the Fourier series of a function $f(x)\in L_1(-\pi,\pi)$ if, and only if, there

THEOREM 5. *If $\alpha(t)$ is a normalized function of bounded variation over $(-\infty, \infty)$, and if $F(x)$ is represented in the form (8.1), then*

$$(9.3)\quad \tfrac{1}{2}\{\alpha(t+0)-\alpha(t-0)\} = (2\pi)^{-\frac{1}{2}} \lim_{s\to\infty} s \int_{-\infty}^{\infty} \frac{e^{ixt}F(x)}{x^2+s^2}\,dx \qquad [-\infty < t < \infty].$$

$$(9.4)\quad \alpha(t)-\alpha(0+)\operatorname{sgn} t = (2\pi)^{-\frac{1}{2}} \lim_{k\to\infty} \int_0^t du \int_{-\infty}^{\infty} \frac{F(x)\,dx}{(1-iux/k)^{k+2}} \qquad \begin{bmatrix} -\infty < t < \infty \\ \operatorname{sgn} 0 = 0 \end{bmatrix}.$$

THEOREM 5′. *If, for $1 \leq p < \infty$, $F(x)$ is the $\mathfrak{M}^q$ Fourier transform of an element $f(t)$ of $L_p(-\infty, \infty)$ $[1/q+1/p=1]$, then*

$$(9.5)\quad F(x) = (2\pi)^{-\frac{1}{2}} \underset{k\to\infty}{\mathfrak{M}^q} \int_{-\infty}^{\infty} \frac{f(t)\,dt}{(1+itx/k)^{k+1}},$$

$$(9.6)\quad f(t) = (2\pi)^{-\frac{1}{2}} \underset{k\to\infty}{\mathfrak{M}^p} \int_{-\infty}^{\infty} \frac{F(x)\,dx}{(1-itx/k)^{k+1}} \qquad [k=1,2,\cdots, k\to\infty].$$

For $q=\infty$, i.e. $p=1$, in (9.5) the limit in mean is to be replaced by the ordinary limit, which is uniform with respect to x in any finite interval.

We observe that (9.6) holds for $p=1$, while, in general, the familiar inversion formula $f(t) = (2\pi)^{-\frac{1}{2}}\mathfrak{M}^p \int e^{itx}F(x)\,dx$ holds for $1 < p < \infty$ only.

From the latter results the necessity of the conditions in Theorems 4(B) and 4′ follows easily, observing that strong convergence in $L_p(-\infty, \infty)$ implies weak convergence and implies therefore also (9.2), and that $g_k[1;t] \equiv 0$ $[k=1,2,\cdots]$. Hence we have only to prove the necessity of (9.2) (A), and the sufficiency of the conditions.

10. Proof of theorem 4(A). First we show that the condition is sufficient.

Since $g_k(x)$ is finite for almost all x, we deduce that $(1+t^2)^{-1}F(t)$

---

is a sequence $n_1, n_2, \cdots$ $[n_j < n_{j+1}]$, or $r_1, r_2, \cdots$ $[r_j \to 1$ as $j\to\infty]$, such that one of the following alternative conditions is satisfied:

(A) $\{\sigma_{n_j}(x)\}$ is a weakly convergent sequence in $L_1(-\pi, \pi)$.

(B) Given $\epsilon > 0$, there is a $\delta = \delta(\epsilon) > 0$ such that, uniformly with respect to j, $\left|\int_E \sigma_{n_j}(x)\,dx\right| < \epsilon$ whenever $m(E) < \delta$ [E a set in $(-\pi, \pi)$].

(C) $\{u(r_j, x)\}$ is a weakly convergent sequence in $L_1(-\pi, \pi)$.

(D) $\left|\int_E u(r_j, x)\,dx\right| < \epsilon$ whenever $m(E) < \delta = \delta(\epsilon)$.

belongs to $L_1(-\infty,\infty)$, taking $k=1$. Now we form the function $\Phi_2(x)$ defined by (1.42), and put $z=-is$ where $s>0$. Then we have

$$(10.1)\quad \frac{d^k}{ds^k}\Phi_2(-is)=(-1)^k i^{k+1}\frac{k!}{2\pi}\int_{-\infty}^{\infty}\frac{F(t)\,dt}{(t+is)^{k+1}}\qquad [k=1,2,\cdots],$$

the differentiation under the integral sign being justified by integrating. Let $h(t)$ be defined in $(0,\infty)$, and let

$$(10.2)\qquad L_{k,u}h=\frac{(-1)^k}{k!}\left(\frac{k}{u}\right)^{k+1}h^{(k)}\left(\frac{k}{u}\right)\qquad \left(\begin{matrix} h^{(k)}(t)=\dfrac{d^k h(t)}{dt^k}\\ k=1,2,\cdots\end{matrix}\right)$$

denote the Post-Widder operator. From (10.1) we have

$$\int_0^{\infty}|L_{k,u}[\Phi_2(-is)]|\,du=\frac{1}{2\pi}\int_0^{\infty}du\,\Big|\int_{-\infty}^{\infty}\frac{F(t)\,dt}{(1-ik^{-1}tu)^{k+1}}\Big|.$$

In consequence of the hypothesis, the right side is bounded with respect to k. This condition, however, together with the existence of derivatives of all orders of $\Phi_2(-is)$ for $0<s<\infty$, is necessary and sufficient [19] for the representation of $\Phi_2(-is)$ in the form

$$\Phi_2(-is)=\int_0^{\infty}e^{-st}d\beta(t)\qquad \left(\begin{matrix} s>0,\\ \int_0^{\infty}|d\beta(t)|<\infty\end{matrix}\right).$$

Both sides are analytic functions of s for $\Re(s)>0$. Taking $s=\sigma+ix$ and $\sigma\to 0$, the left side tends, for almost all x in $(-\infty,\infty)$, to the limit-function $F_2(x)$ (see Theorem 2), while the right side tends to $\int_0^{\infty}e^{-ixt}d\beta(t)$ for all x. Hence

$$F_2(x)\equiv\int_0^{\infty}e^{-ixt}d\beta(t)$$

for almost all x. In a similar way we deduce that

$$\begin{matrix}\phi_1(is)=\int_0^{\infty}e^{-st}d\gamma(t)\,;\quad F_1(-x)\equiv\int_0^{\infty}e^{-ixt}d\gamma(t),\\ F_1(x)\equiv\int_{-\infty}^{0}e^{-ixt}d\tilde{\gamma}(t)\end{matrix}\qquad \left(\begin{matrix}\Re(s)>0;\\ \int_0^{\infty}|d\gamma(t)|<\infty,\\ \tilde{\gamma}(t)=-\gamma(-t)\end{matrix}\right).$$

Observing that $F(x)=F_1(x)+F_2(x)$, it is easy to complete the proof.

To prove the necessity of the condition, we compare the resolution of $F(x)$

---

[19] D. V. Widder, *loc. cit.*, Ch. 7, Theorem 12a.

into two components, based on the representation (8.1), with that given by Theorems 2′, 2″ (a). Thus we have

$$\Phi_2(-is) = (2\pi)^{-\frac{1}{2}} \int_0^\infty e^{-ts} d\alpha(t) + A,$$

$$\Phi_1(is) = (2\pi)^{-\frac{1}{2}} \int_0^\infty e^{-ts} d\{-\alpha(-t)\} - A.$$

Applying the Widder theorem, we deduce (9.2 A), which completes the proof.

11. The case (B) is treated in a similar way. We need some lemmas.

LEMMA 5. *Let $F(t)$ satisfy* (9.2) (A) *or* (B). *Then both $\Phi_1(is)$ and $\Phi_2(-is)$ tend to finite limits as $s \to \infty$.*

LEMMA 6. *Let $1 < p < \infty$, $1/p + 1/q = 1$, and let $H(-x)$ be the $\mathfrak{M}^q$ Fourier transform of a function $h(t) \in L_p(0, \infty)$ [i. e., $h(t) \equiv 0$ for $t < 0$]. Then $H(x)$ is the limit-function of an element $H(z)$ of $\mathfrak{H}_q$, defining $H(z)$ by*

$$(2\pi)^{-\frac{1}{2}} H(is) = \int_0^\infty e^{-st} h(t)\, dt \qquad [\Re s > 0]. \tag{11.1}$$

LEMMA 7.[20] *If* (i) *$f(s)$ has derivatives of all orders for $0 < s < \infty$ and* (ii) *$f(\infty) = 0$ and* (iii), *for some finite $p > 1$,*

$$\int_0^\infty | L_{k,u} f |^p du < M \qquad [k = 1, 2, \cdots], \tag{11.2}$$

then $f(s)$ is representable in the form

$$f(s) = \int_0^\infty e^{-st} h(t)\, dt \qquad [0 < s < \infty],$$

where $h(t) \in L_p(0, \infty)$. The conditions are necessary.

Proof of Lemma 5. We need only show that the function

$$\chi(s) = 2\pi \frac{d}{ds} \Phi_1(is) = \int_{-\infty}^\infty \frac{F(t)\, dt}{(t - is)^2}$$

belongs to $L_1(1, \infty)$. Certainly $s^{2-2/p}\chi(s) \in L_p(0, \infty)$ $[1 \leq p \leq 2]$; for

$$\int_0^\infty s^{2p-2} | \chi(s) |^p ds = \int_{-\infty}^0 du \left| \int_{-\infty}^\infty \frac{F(t)\, dt}{(1 - itu)^2} \right|^p = \int_{-\infty}^\infty | g_1(u) |^p du < M.$$

Using Hölder's inequality if $p > 1$, we arrive at the required result.

[20] D. V. Widder, Ch. 7, Theorem 15a.

Lemma 6 follows from known results, due to E. Hille and J. D. Tamarkin,[21] and to G. Doetsch.[22]

Proof of the sufficiency of the condition (*case* B). By Lemma 5, $\Phi_2(-i\infty)$ exists and is finite. The function $\Phi_2(-is)-\Phi_2(-i\infty)$ satisfies the hypotheses of Lemma 7. Hence there is a function $h_2(t)\,\epsilon\, L_p(0,\infty)$ such that

$$\Phi_2(-is)=\int_0^\infty e^{-st}h_2(t)\,dt+\Phi_2(-i\infty) \qquad [s>0].$$

By Theorem 2 and Lemma 6, therefore, we have $F_2(x)=H(-x)+\Phi_2(-i\infty)$, where $H(x)$ is defined by

$$H(-x)=\mathfrak{M}^q\int_0^\infty e^{-ixt}h_2(t)\,dt.$$

Similarly

$$F_1(x)=\mathfrak{M}^q\int_{-\infty}^0 h_1(t)e^{-ixt}dt+\Phi_1(i\infty) \qquad [h_1(t)\,\epsilon\, L_p(-\infty,0)].$$

Using Theorem 2, we arrive at the required result.

12. Proof of Theorem 4′. The result for $p>1$ is now evident. We are left to show that for the case $p=1$ the condition is sufficient. We need a result which can be considered as a generalisation of a Widder theorem.[23]

LEMMA 8. *If* (i) $f(s)$ *has derivatives of all orders for* $0<s<\infty$ *and* (ii) $f(\infty)=0$ *and* (iii) $\{L_{k,u}f\}$ $[k=1,2,\cdots]$ *is a weakly convergent sequence in* $L_1(0,\infty)$, *then* $f(s)$ *is representable in the form* $f(s)=\int_0^\infty e^{-st}g(t)\,dt$, *where* $0<s<\infty$ *and* $g(t)\,\epsilon\, L_1(0,\infty)$. *The conditions are necessary.*

In the Widder theorem, $\{L_{k,u}f\}$ is required to be a strongly convergent sequence in $L_1(0,\infty)$. Therefore we need only show that the conditions are sufficient. If $g_0(u)$ is the element of $L_1(0,\infty)$ to which $\{L_{k,u}f\}$ converges weakly, and if $f_0(s)=\int_0^\infty e^{-st}g_0(t)\,dt$, then, by the Widder theorem, the $L_{k,u}f_0$ converge strongly in $L_1(0,\infty)$; in fact they converge to $g_0(t)$. Hence they converge to $g_0(t)$ also weakly; therefore the sequence $\{L_{k,u}[f-f_0]\}$ converges weakly as well, and so

[21] *Loc. cit.*, Lemma 4. 2.

[22] *Mathematische Zeitschrift*, vol. 42 (1937), pp. 263-286.

[23] D. V. Widder, Ch. 7, Theorem 17a.

$$\int_0^\infty | L_{k,u}[f - f_0]| \, du < M \qquad [k = 1, 2, \cdots].$$

Consequently $f(s) - f_0(s)$ is representable in the form [24]

$$f(s) - f_0(s) = h(s) = \int_0^\infty e^{-st} d\alpha(t) \qquad [s > 0, \int_0^\infty | d\alpha(t)| < \infty].$$

But the $L_{k,u}h$ converge to the null-element, and we have [25]

$$\alpha(t) = \lim_{j\to\infty} \int_0^t L_{n_j,u}h \, du + \beta(t) \qquad \left[\beta(t) = \begin{matrix} -h(\infty) \\ 0 \end{matrix} \text{ for } \begin{matrix} t = 0 \\ t > 0 \end{matrix}\right],$$

where $\{n_j\}$ is some sequence tending to infinity. Hence $\alpha(t) = \beta(t)$. Since $h(\infty) = f(\infty) - f_0(\infty) = 0 - 0 = 0$, we have $\alpha(t) = 0$ $[0 \leqq t < \infty]$ and, therefore, $f(s) - f(s_0) = 0$, which proves the lemma.

Using this result, the proof can now be completed by arguments advanced in the preceding section.

13. The Stieltjes transformation

$$(13.1) \qquad Tf = T[f(t); x] = \frac{1}{\pi} \int_0^\infty \frac{f(t)\,dt}{t + x} \qquad [0 < x < \infty].$$

A solution of the Stieltjes problem in terms of functions of a real variable has been given by D. V. Widder, also for the case $\tilde{T}f = \pi^{-1} \int_0^\infty (t + x)^{-1} df(t)$. Here we shall deal with (13.1) only, using analytic functions, and shall generalise a result known for the case $f(t) \epsilon L_2(0, \infty)$.[26]

THEOREM 6. (A) *If* (i) *both* $g(z)$ *and* $g(-z)$ *belong to* $\mathfrak{H}_p$ $[1 \leqq p < \infty]$, *and if* (ii) $g(z)$ *is analytic for* $0 < x < \infty$, *then the equation* $g(x) = Tf$ *has a unique solution* $f(t) \epsilon L_p(0, \infty)$. *The conditions* (ii) *and, for* $1 < p < \infty$, (i) *are necessary.*

(B) *If* (i) *both* $(z + i)^{-1}g(z)$ *and* $(z + i)^{-1}g(-z)$ *belong to* $\mathfrak{H}_1$, *and* (ii) $g(z)$ *is analytic for* $0 < x < \infty$, *then* $g(x) = Tf$ *has a unique solution* $f(t)$ *such that* $(1 + t)^{-1}f(t)$ *is integrable on* $(0, \infty)$.

---

[24] D. V. Widder, Ch. 7, Theorem 12a.

[25] D. V. Widder, p. 307.

[26] E. C. Titchmarsh, *loc. cit.*, 11. 8. It can be shown that the Titchmarsh condition is equivalent to those of Theorem 6 by the following result: $F(z)$ belongs to $\mathfrak{H}_p$ $[0 < p < \infty]$ if and only if $F(z)$ is analytic for $y > 0$ and if, uniformly for $0 < \theta < \pi$,

$$\int_0^\infty | F(re^{i\theta})|^p dr < M^p.$$

In both cases the solution is.

$$(13.2)\quad f(t) = \frac{1}{2i} \lim_{y\to 0} \{g(-t-iy) - g(-t+iy)\} \left[\begin{matrix} y > 0, \\ -\infty < t < \infty \end{matrix}\right].$$

Proof of the uniqueness of the solution. It will suffice to show that the conditions $(1+t)^{-1}f_j(t) \in L_1(0,\infty)$ and $Tf_j(t) = g(x)$ $[j=1,2]$ imply $f_1(t) \equiv f_2(t)$ $[0 < t < \infty]$. Let $f_1(t) - f_2(t) = f(t)$; $\gamma(z) = T[f(t);z]$. Then $\gamma(z)$ is analytic for $|\arg z| < \pi$, while $\gamma(x) = Tf_1 - Tf_2 = 0$ for $0 < x < \infty$. Hence $\gamma(z)$ vanishes identically. Taking $F(t) = 0$ for $t < 0$, $F(t) = f(t)$ for $t > 0$, and using (1.41) and Theorem 2, we have $F(t) = F_1(t) + F_2(t)$, and $F_j(z) = (-1)^{j-1}(2i)^{-1}\gamma(-z)$ for $y > 0$ $(j=1)$ or $y < 0$ $(j=2)$, respectively. Thus $F_j(z)$ vanishes identically; so does, therefore, $F(t)$ and $f_1(t) - f_2(t) = f(t)$.

14. Proof of part (A) of the theorem. The necessity of (ii) is evident. So is that of (i) for $1 < p < \infty$ by (3C), taking $F(t) = f(t)$ for $t > 0, = 0$ for $t < 0$. If, for $p = 1$, we require further that $f(t) \in L_1(0,\infty)$ and that $\mathfrak{H}f \in L_1(-\infty,\infty)$, then the necessity of (i) follows from Theorem 2′.

The function $f(t)$, defined by (13.2), has the properties required. Since $g(z)$ is analytic for $0 < x < \infty$, $f(t)$ vanishes for $t < 0$. In consequence of (i), $f(t)$ belongs to $L_p(0,\infty)$. Taking

$$\psi_1(t) = \lim_{y\to 0} g(-t-iy), \qquad \psi_2(t) = \lim_{y\to 0} g(-t+iy) \qquad [y > 0],$$

we have [27]

$$\mathfrak{H}\psi_1 = i\psi_1, \qquad \mathfrak{H}\psi_2 = -i\psi_2$$

as $\psi_1(t)$ and $\psi_2(-t)$ are limit functions of elements of $\mathfrak{H}_p$. Hence

$$f(t) = \frac{1}{2i}(\psi_1(t) - \psi_2(t))$$

$$\mathfrak{H}[f(t);x] = \tfrac{1}{2}(\psi_1(x) + \psi_2(x)) \in L_p(-\infty,\infty).$$

In the latter equation the term on the left is equal to $T[f(t);-x]$ for $x < 0$ and the term on the right to $g(-x)$ which proves the case (A) of the theorem. The case (B) is treated in a similar way, using Lemma 10 (**15**).

15. Generalisation of known results in $\mathfrak{H}_p$ and on Hilbert's operator.

LEMMA 9. *If* $(z+i)^{-1}F(z) \in \mathfrak{H}_1$ *or* $(z+i)^{-2}F(z) \in \mathfrak{H}_1$, *respectively, then*

[27] H. Kober, *Bulletin of the American Mathematical Society*, vol. 48 (1942), *loc. cit.*, Lemma 4.

$$F(z)=\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{F(t)\,dt}{t-z} \quad or \quad F(z)=\frac{1}{2\pi i}\int_{-\infty}^{\infty}F(t)\left(\frac{1}{t-z}-\frac{1}{t+i}\right)dt$$

for $y>0$, while the integrals vanish identically for $y<0$.

LEMMA 10. *The function $(t+i)^{-j}F(t)$ $(j=1,2)$ is the limit-function of an element of $\mathfrak{H}_1$ if, and only if, $(1+|t|)^{-j}F(t)\in L_1(-\infty,\infty)$ and $\mathfrak{H}F=iF$ or $\mathfrak{K}F=i(F-a_0)$, respectively, where $a_0=\pi^{-1}\int_{-\infty}^{\infty}(1+t^2)^{-1}F(t)\,dt$.*

LEMMA 11. *If* (i) $(1+|t|)^{-j}F(t)$ $(j=1,2)$ *and* (ii) $(1+|x|)^{-1}\mathfrak{H}F$ *or* $(1+x^2)^{-1}\mathfrak{K}F$, *respectively, belong to* $L_1(-\infty,\infty)$, *then*

$$\mathfrak{H}^2F=-F(x)\ [j=1] \quad or \quad \mathfrak{K}^2F=-F(x)+a_0\ [j=2].$$

If $(z+i)^{-1}F(z)\in\mathfrak{H}_1$, then the statement concerned in Lemma 9 is deduced from the representation of a function $g(z)\in\mathfrak{H}_1$ by its proper Cauchy integral.[28] We take $g(z)=(z+i)^{-1}F(z)$ and observe that $g(z)\in\mathfrak{H}_1$ implies that $\int_{-\infty}^{\infty}g(t)\,dt=0$. The other statement is reduced to the previous case by taking $(z+i)^{-1}F(z)=\tilde{F}(z)$.

The proof of Lemma 10 is left to the reader.

To deduce Lemma 11 for $j=2$, we resolve $F(t)$ into its components according to Theorem 2 and (4.4). Thus we have

$$F(t)=F_1(t)+F_2(t), \tag{15.1}$$

$$\mathfrak{K}F=iF_1(t)-iF_2(t). \tag{15.2}$$

By Theorem 2‴. (4), $F_j(x)$ $(j=1,2)$ is the limit-function of a function $\Phi_j(z)$ such that $(z+i)^{-2}\Phi_1(z)$ or $(z+i)^{-2}\Phi(-z)$ belongs to $\mathfrak{H}_1$. By Lemma 10, therefore, we have

$$\mathfrak{K}F_j=(-1)^{j-1}i\left(F_j(x)-\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{F_j(t)\,dt}{t^2+1}\right).$$

Applying the operator $\mathfrak{K}$ to (15.2), we have

$$\mathfrak{K}^2F=-F_1(x)+\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{F_1(t)\,dt}{t^2+1}-F_2(x)+\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{F_2(t)\,dt}{t^2+1}=-F(x)+a_0.$$

In a similar way we deduce the other assertion of the lemma.

THE UNIVERSITY,
BIRMINGHAM, ENGLAND.

[28] Hille-Tamarkin, *loc. cit.*, Theorem 2.1 (ii).

HADAMARD'S FORMULA AND VARIATION OF DOMAIN-FUNCTIONS.*

By MENAHEM SCHIFFER.

1. Hadamard's differential equation for Green's function.

1. The present paper deals with applications of functional analysis to the theory of the logarithmic potential and analytic functions. We consider domains D in the complex x-plane bounded by a finite number of proper continua $C_\nu (\nu = 1, \cdots, n)$. If x and y are points in D, Green's function $g(x;y)$ of D is defined in the following way:

a. $g(x;y)$ is a harmonic function of x throughout D, the point $x = y$ excepted; there however $g(x;y) + \log |x - y|$ is harmonic.

b. If x converges to a boundary continuum C_ν, $g(x;y)$ converges to zero

The existence of $g(x;y)$ is assured for every domain D; $g(x;y)$ is harmonic in y also and satisfies the symmetry condition $g(x;y) = g(y;x)$ Its importance for Dirichlet's problem is well known. If D is simply connected, $g(x;y)$ is related also to the problem of mapping D onto the exterior E of the unit circle. For, let $p(x;y)$ be an analytic function of x in D, such that $g(x;y) = R\{p(x;y)\}$; then $\phi(x;y) = \exp\{p(x;y)\}$ maps D conformally on E so that the point y goes into infinity.

Using the theory of orthogonal functions we may determine the Green's function in the form of an infinite series for any domain whose boundary satisfies certain general conditions. (See Bergman[1] 1, 2.). On the other hand consideration of the rôle of Green's formula in the theory of functions and potential theory indicates the desirability of obtaining various representations of it and in particular those which show the dependence of this formula on the domain in which it is defined.

There are however only a few elementary domains D with an explicitly known Green's function. Nevertheless it is possible to calculate Green's function (approximately) for domains which are sufficiently near such elementary domains. This is done by means of Hadamard's well known variation formula (see Hadamard 2) which is applicable to every domain D bounded by analytic

* Received November 15, 1945.

[1] See Bibliography, p. 448.

curves C_ν. Let every point of the boundary $C = \sum_{\nu=1}^{n} C_\nu$ be defined by a parameter s which measures the lengths of the curves successively and runs, therefore, from 0 to l (l = sum of the lengths of all C_ν). Let $\delta n(s) = \epsilon\nu(s)$ be a continuous function of s which determines the normal displacement of each boundary point $z(s)$ of the original domain D. $\delta n(s)$ is taken as positive if the displacement is in the direction of the outer normal with respect to D. In this way, we define a new domain D^* with a new Green's function $g^*(x;y)$. According to Hadamard, we have

$$(1) \qquad g^*(x;y) = g(x;y) + \frac{\epsilon}{2\pi}\int_C \frac{\partial g(z;x)}{\partial n_z}\,\frac{\partial g(z;y)}{\partial n_z}\,\nu(s_z)\,ds_z + o(\epsilon).$$

Here, $\frac{\partial}{\partial n_z}[g(z;x)]$ denotes the derivative of $g(z;x)$ in the direction of the outward normal and $o(\epsilon)$, as usual, a term satisfying the condition $\lim_{\epsilon=0}\frac{1}{\epsilon}o(\epsilon) = 0$. Using the notations of functional calculus (Volterra 1), we may give to (1) the form

$$(1') \qquad \delta g(x;y) = \frac{1}{2\pi}\int_C \frac{\partial g(z;x)}{\partial n}\,\frac{\partial g(z;y)}{\partial n}\,\delta n\,ds.$$

Many properties of $g(x;y)$ and the univalent mapping functions connected with it may be derived from (1′) (see Lévy, Julia, Biernacki). This formula, however, loses its meaning if the initial domain D is not bounded by smooth curves; on the other hand, in this case a well defined Green's function also exists. Thus (1′) is not applicable to extremum problems concerning Green's function; for one can not be sure that the domain D, belonging to the extremal function, satisfies the suppositions of Hadamard's formula.

2. We shall now transform formula (1′) into such a form that it may be applied to the most general domain D. We use the artifice of specializing the variation δn and, by means of partial integration, expressing $\delta g(x;y)$ by values of $g(x;y)$ and its derivatives from the *interior* of D. For this purpose, we choose a fixed point z_0 in D and consider the representation

$$(2) \qquad x^* = x + \frac{e^{2i\phi}\rho^2}{x - z_0}, \quad \rho > 0, \qquad 0 \leq \phi < 2\pi$$

of the x-plane. It transforms the circumference $|x - z_0| = \rho$ into the segment $< -2\rho e^{i\phi}, +2\rho e^{i\phi} >$ and is univalent in its exterior $|x - z_0| > \rho$. For ρ sufficiently small, this representation is univalent on all curves C_ν and transforms them in a one-to-one manner into neighboring curves C^*_ν which

enclose a new domain D^* of the x-plane. Let $g^*(x;y)$ denote the Green's function of D^*; we shall compute it from $g(x;y)$ by aid of (1′).

We remark that for $z \subset C_\nu$, $y \subset D$

$$(3) \qquad g(z^*;y) = \frac{\partial g(z;y)}{\partial n}\,\delta n + o(\rho^2);$$

here δn is the normal shift of C_ν at the point z, caused by the variation (2). δn is of order ρ^2. Hence (1) has the form

$$(4) \qquad g^*(x;y) = g(x;y) + \frac{1}{2\pi}\int_C \frac{\partial g(z;x)}{\partial n}\, g(z^*(z);y)\,ds + o(\rho^2).$$

Let y_1 be the point which is transformed by (2) into y; since $g(z;y_1)$ vanishes on the boundary of D, we may write instead of (4)

$$(4') \qquad \delta g(x;y) = \frac{1}{2\pi}\int_C \frac{\partial g(z;x)}{\partial n}\,[g(z^*;y) - g(z;y_1)]\,ds + o(\rho^2).$$

Now, $g(z^*;y) - g(z;y_1)$ is harmonic in the domain D_0, obtained from D by discarding the interior of the circumference $K_\rho \equiv (|x - z_0| = \rho)$. Hence, Green's formula may be appplied to D_0, yielding

$$(5) \qquad \delta g(x;y) = \frac{1}{2\pi}\int_{K\rho}\Big\{\frac{\partial g(z;x)}{\partial n}[g(z^*;y) - g(z;y_1)]$$
$$- g(z;x)\,\frac{\partial}{\partial n}\,[g(z^*;y) - g(z;y_1)]\Big\}\,ds + g(x;y_1) - g(x^*;y) + o(\rho^2)$$

where the normal derivative is to be taken in the direction of increasing ρ. We suppose x, y, y_1 to lie in the exterior of K_ρ and get, therefore, from Green's formula

$$(5') \qquad \frac{1}{2\pi}\int_{K\rho}\Big\{\frac{\partial g(z;x)}{\partial n}\,g(z;y_1) - g(z;x)\,\frac{\partial g(z;y_1)}{\partial n}\Big\}\,ds = 0.$$

Thus, there remains

$$(6) \qquad \delta g(x;y) = \frac{1}{2\pi}\int_{K\rho}\Big\{\frac{\partial g(z;x)}{\partial n}\,g(z^*;y) - g(z;x)\,\frac{\partial g(z^*;y)}{\partial n}\Big\}\,ds$$
$$+ g(x;y_1) - g(x^*;y) + o(\rho^2)$$

which expresses the variation of $g(x;y)$ by means of the values of $g(x;y)$ on the circumference K_ρ, interior to D, only. This formula may be simplified by means of a development in series. We introduce an auxiliary function $p(x;y)$ analytic in x and satisfying

$$(7) \qquad g(x;y) = R\{p(x;y)\}.$$

$p(x;y)$ is not necessarily uniform in D and contains an arbitrary additive imaginary constant depending on y. On K_ρ we have $z = z_0 + \rho e^{i\tau}$ so that, on applying Taylor's theorem to $p(x;y)$ and by (7)

$$(8) \qquad g(z^*;y) = g(z_0;y) + R\{\rho(e^{i\tau} + e^{i(2\phi-\tau)})p'(z_0;y)\} + O(\rho^2),$$

$$(8') \qquad \frac{\partial}{\partial n} g(z^*;y) = R\{(e^{i\tau} - e^{i(2\phi-\tau)})p'(z_0;y)\} + O(\rho),$$

$$(8'') \qquad g(z;x) = g(z_0;x) + R\{\rho(e^{i\tau}p'(z_0;x)\} + O(\rho^2),$$

$$(8''') \qquad \frac{\partial}{\partial n} g(z;x) = R\{e^{i\tau}p'(z_0;x)\} + O(\rho),$$

$$(8^{\mathrm{IV}}) \qquad g(x;y_1) = g(x;y) - R\left\{\frac{e^{2i\phi}\rho^2}{y-z_0} p'(y;x)\right\} + o(\rho^2),$$

$$(8^{\mathrm{V}}) \qquad g(x^*;y) = g(x;y) + R\left\{\frac{e^{2i\phi}\rho^2}{x-z_0} p'(x;y)\right\} + o(\rho^2).$$

The dash denotes differentiation of $p(x;y)$ with respect to its first argument and $O(\epsilon)$ a term satisfying the condition that $(1/\epsilon)O(\epsilon)$ remains bounded if $\epsilon \to 0$.

Introducing formulae (8) into (6) we obtain

$$(9) \qquad \delta g(x;y) = -R\left\{ e^{2i\phi}\rho^2 \left[\frac{p'(x;y)}{x-z_0} + \frac{p'(y;x)}{y-z_0}\right]\right\} + \frac{1}{2\pi}\int_0^{2\pi} 2R\{e^{i\tau}p'(z_0;x)\}\cdot R\{e^{i(2\phi-\tau)}p'(z_0;y)\}\rho^2 d\tau + o(\rho^2)$$

and by an easy transformation

$$(10) \qquad g^*(x;y) = g(x;y) + R\left\{ e^{2i\phi}\rho^2 \left[p'(z_0;x)p'(z_0;y) - \frac{p'(x;y)}{x-z_0} - \frac{p'(y;x)}{y-z_0}\right]\right\} + o(\rho^2).$$

In virtue of the identity

$$(10') \quad g^*(x^*;y^*) = g(x;y) + R\left\{ e^{2i\phi}\rho^2 \left[\frac{p'(x;y)}{x-z_0} + \frac{p'(y;x)}{y-z_0}\right]\right\} + o(\rho^2),$$

we may put (10) into the simple form (Schiffer 3)

$$(11) \qquad g^*(x^*;y^*) = g(x;y) + R\{e^{2i\phi}\rho^2 p'(z_0;x)p'(z_0;y)\} + o(\rho^2)$$

giving the law of variation of $g(x;y)$ under the particular transformation (2) of D.

(11) was derived from (1) by partial integration. The latter formula is only valid for analytically bounded domains; the same holds, therefore, for (11). But the most general domain D may be approximated arbitrarily

by analytically bounded domains; Green's function of the approximating domain with all its derivatives tends to the corresponding terms of the given domain, uniformly in every interior part. $o(\rho^2)$ depends only upon $g(x;y)$ and its derivatives at interior points of D; hence it can be estimated uniformly. Thus the validity of (11) can be derived for the most general domains D. This fact proves (11) to be superior to (1′), and numerous applications of this formula are possible in extremum problems, concerning the theory of the logarithmic potential and conformal representation, as will be seen below.

Another application of our method is as follows: In the paper (Schiffer 4) the relation between Green's function and the kernal function (see Bergman 1, §VII) is derived. The methods used in the present paper can be employed in investigating the kernal function and lead to new results concerning the behavior of this function in simply- and multiply-connected domains.

2. Further solutions of the variational equation (11).

1. The differential equation (1′) for Green's function admits as solution also other domain functions depending on two variables x, y (see Lévy). If the values of such a function are known for an elementary initial domain, say a circle, it can be calculated in principle for *every* other domain by means of (1′). But it will, in general, be difficult to characterize such a function geometrically, independently of the definition by means of a variational equation. On the other hand, we shall define now an important family of domain functions which occur in the theory of conformal representation of multiply connected domains and which satisfy (11).

In the theory of Green's function the boundary condition plays a decisive part. We shall define a more general type of boundary condition which renders the same service and will be called henceforth the type N. It is characterized by the following property:

Let C_ν $(\nu = 1, \cdots, n)$ be a system of smooth curves enclosing a domain D and let $z^*(z)$ be a conformal representation in the neighborhood of all the C_ν, transforming them into a new system of curves C^*_ν which encloses the domain D^*. A boundary condition is said to be of type N, if it assures in this case the following two facts: $\phi(x)$ satisfying this condition with respect to D, $\psi^*(x)$ with respect to D^*, implies that

a. $\phi(x)$ has continuous derivatives on C_ν, and $\psi^*(x)$ on C^*_ν.

b. The following relation is always valid:

$$\int_C \{\phi(z) \frac{\partial}{\partial n_z} \psi^*(z^*(z)) - \psi^*(z^*(z)) \frac{\partial}{\partial n_z} \phi(z)\} ds_z = 0. \tag{12}$$

Next, domain functions $\gamma(x;y)$ of D with the following properties will be considered:

a. $\gamma(x;y)$ is, for $y \subset D$ fixed, a harmonic function of $x \subset D$, the point $x = y$ excepted.

b. In the neighborhood of $x = y$, the expression $\gamma(x;y) + \log|x - y|$ is bounded.

c. The function $\gamma(x;y)$ depends continuously on D, uniformly with respect to x.

d. $\gamma(x;y)$ satisfies as function of x a boundary condition of type N.

We assert that on these conditions $\gamma(x;y)$ has a variation formula (11).

First, it is obvious that $\gamma(x;y)$ is symmetric with respect to both its arguments. For, on one hand, Green's identity yields in the case of a smooth boundary C

$$(13) \qquad \frac{1}{2\pi}\int_C \{\gamma(z;x)\frac{\partial}{\partial n}\gamma(z;y) - \gamma(z;y)\frac{\partial}{\partial n}\gamma(z;x)\}ds = \gamma(x;y) - \gamma(y;x),$$

while, on the other hand, the boundary condition of $\gamma(x;y)$ ensures, by virtue of (12), that this integral vanishes; we have only to put $\phi(z) = \gamma(z;x)$, $z^*(z) = z$, $\psi^*(z) = \gamma(z;y)$. Hence, we have proved the symmetry of $\gamma(x;y)$, for domains D, bounded by smooth curves. But in view of the continuity of $\gamma(x;y)$ in dependence on the domain, this establishes the same property for domains with general boundary.

Let D denote a domain with smooth boundary curves C_ν; by means of the variation (2) it is transformed into D^* with the corresponding domain function $\gamma^*(x;y)$. As in **1**, **2**, we construct the domain D_0 by discarding from D the interior of the circumference $K_\rho \equiv (|x - z_0| = \rho)$. In D_0, the function $d(x;y) = \gamma^*(x^*(x);y^*(y)) - \gamma(x;y)$ is harmonic in both its arguments, the logarithmic pole $x = y$ being cancelled by subtraction. Hence, Green's identity yields easily

$$(14) \qquad \frac{1}{2\pi}\int_{C+K_\rho}\{\gamma(z;x)\frac{\partial}{\partial n}d(z;y) - d(z;y)\frac{\partial}{\partial n}\gamma(z;x)\}ds = d(x;y) = \gamma^*(x^*;y^*) - \gamma(x;y).$$

But γ and γ^* satisfy a boundary condition of type N which shows that the integral over C vanishes. Combining this fact with Green's identity

$$(15) \qquad \frac{1}{2\pi}\int_{K_\rho}\{\gamma(z;x)\frac{\partial}{\partial n}\gamma(z;y) - \gamma(z;y)\frac{\partial}{\partial n}\gamma(z;x)\}ds = 0$$

which holds for x and y in D_0, we get finally from (14)

$$(15') \qquad \gamma^*(x^*;y^*) = \gamma(x;y) + \frac{1}{2\pi}\int_{K\rho} \{\gamma(z;x)\frac{\partial}{\partial n_z}\gamma^*(z^*;y^*) - \gamma^*(z^*;y^*)\frac{\partial}{\partial n_z}\gamma(z;x)\}ds_z$$

where the normal derivative is to be taken in the outward direction with respect to D_0. If we want to give to n the direction of increasing radius ρ, as is usual, we get

$$(15'') \qquad \gamma^*(x^*;y^*) = \gamma(x;y) + \frac{1}{2\pi}\int_{K\rho} \{\gamma^*(z^*;y^*)\frac{\partial}{\partial n_z}\gamma(z;x) - \gamma(z;x)\frac{\partial}{\partial n_z}\gamma^*(z^*;y^*)\}ds_z.$$

Thus, we have expressed the variation of $\gamma(x;y)$ by an integral taken along a circumference interior to D as we did in (6) with respect to $g(x;y)$. We may perform now the same formal transformations and developments into series as in **1**, 2; these hold for $\gamma(x;y)$ too, since we have used there only the symmetry and harmonicity of $g(x;y)$. We introduce a function $\phi(x;y)$, analytic for given $y \subset D$ with respect to its first argument $x \subset D$ and satisfying

$$(16) \qquad \gamma(x;y) = R\{\phi(x;y)\}.$$

In general, $\phi(x;y)$ will not be uniform in D; it possesses additive (imaginary) moduli with respect to circuits around the C_ν. We may perform on it the same operations as we did above on $p(x;y)$ and we get, in complete analogy to (11),

$$(17) \qquad \gamma^*(x^*;y^*) = \gamma(x;y) + R\{e^{2i\phi}\rho^2\phi'(z_0;x)\phi'(z_0;y)\} + o(\rho^2).$$

This is the law of variation of $\gamma(x;y)$ under the transformation (2) of a domain D with a smooth boundary. But in view of the uniform continuity of $\gamma(x;y)$ this formula remains valid for domains with general boundary.

2. Our next task is to point out definite examples of domain-functions of the above type. It is well known that every domain D can be mapped in an infinity of ways on the exterior of a circle, cut along concentric circular slits (Koebe 1, Grötzsch 1). This canonical representation is in many respects a natural generalization of the representation of a simply connected domain on the exterior of a circle.

If $f(x;y)$ is a univalent function of x in a simply-connected domain D and maps D on the exterior E of the unit circle such that $y \subset D$ corresponds to infinity, then the Green's function of D is given by

$$(18) \qquad g(x;y) = \log|f(x;y)|.$$

In the case of a multiply connected domain, however, there exists for every $m = 1, 2, \cdots, n$ a function $f_m(x; y)$, mapping D on the exterior of the unit circle, slit along $n-1$ concentric circular arcs, such that C_m corresponds to the unit circle and y to infinity. Thus there exist n domain functions

$$\gamma_m(x; y) = \log |f_m(x; y)| \tag{19}$$

which are a generalization of Green's function for a simply connected domain. We want to show that each $\gamma_m(x; y)$ has a variation formula (17), in analogy to $g(x; y)$. To prove this, it suffices, in view of 1, to show that the boundary conditions for $\gamma_m(x; y)$ are of type N.

Now, there are two characteristic properties of $\gamma_m(x; y)$: (a) $\gamma_m(z; y)$ is zero for z on C_m and constant on each C_ν. (b) $f_m(x; y)$ being uniform in D, the conjugate function of $\gamma_m(x; y)$, i. e., $I\{\log f_m(x; y)\}$, does not change when x describes a circuit around C_ν $(\nu \neq m)$ which means, in the case of a smooth C_ν,

$$\int_{C_\nu} \frac{\partial}{\partial n_z} \gamma_m(z; y)\, ds_z = 0 \qquad (\nu \neq m). \tag{20}$$

These two facts represent just a boundary condition of type N. For, if any $\phi(z)$ possesses the properties (a) and (b) on the system of smooth curves C_ν and any $\psi^*(z)$ on the corresponding system C^*_ν, then we have for each curve C_ν

$$\int_{C_\nu} \{\phi(z) \frac{\partial}{\partial n_z} \psi^*(z^*(z)) - \psi^*(z^*(z)) \frac{\partial}{\partial n_z} \phi(z)\} ds_z = 0. \tag{12'}$$

In fact, $\phi(z)$ and $\psi^*(z^*(z))$ vanish on C_m; on each other C_ν both functions are constant and may be taken out of integration. Thus, there remain periods of type (20) which vanish. Hence, the boundary conditions considered are of type N and $\gamma_m(x; y)$ satisfies, therefore, a variational equation (17).

3. We mention now another type of canonical representation which may again be considered as a generalization of the representation onto the exterior E of the unit circle in the case of a simply connected domain. It is the representation of D by means of a function $h_m(x; y)$ onto E, slit along $n-1$ radial segments, such that C_m corresponds to the circle and y to infinity (Grötzsch 1, Rengel 1, Koebe 1). Consider now the domain function

$$\kappa_m(x; y) = \log |h_m(x; y)|. \tag{21}$$

This function too possesses a variation formula (17), since it has boundary conditions of type N. To show this, we remark the following two facts:

(a) $\kappa_m(x;y)$ vanishes on C_m. (b) On every C_ν $(\nu \neq m)$ the conjugate function of $\kappa_m(x;y)$, i. e., $I\{\log h_m(x;y)\}$ is constant. Hence, using Cauchy-Riemann's differential equations, we have for $z \subset C_\nu$ $(\nu \neq m)$

$$(22) \qquad \frac{\partial}{\partial n_z} \kappa_m(z;y) = -\frac{\partial}{\partial s_z} I\{\log h_m(z;y)\} = 0.$$

These conditions ensure the validity of relations of type (12) for every smooth C_ν; for, $\kappa^*_m(x^*;y^*)$, too, has a constant conjugate function if z remains on C_ν $(\nu \neq m)$. The boundary conditions (a) and (b) are, therefore, of type N and $\kappa_m(x;y)$ satisfies a variational equation (17).

We may consider other canonical representations of D, by means of univalent functions $l_m(x;y)$ which map D onto E slit by concentric circular arcs and radial segments, such that $y \subset D$ corresponds to infinity, C_m to the circle and the other C_ν to either of the slits. It follows easily that the domain function

$$(23) \qquad \lambda_m(x;y) = \log |l_m(x;y)|$$

is of the type $\gamma(x;y)$ characterized in 1. Hence it varies too according to (17).

Thus, we have found various domain functions with the same law of variation as Green's function. The function $p(x;y)$ in (11) can only with difficulty be interpreted geometrically, whereas the functions $\phi(x;y)$, used in (17) for the variation of our domain functions, are the logarithms of univalent functions yielding canonical representations. The next paragraphs will show the advantages which evolve from this fact.

4. Next, we transform (17) in such a way that its relation to Hadamard's formula is exposed. From (17) we obtain easily in virtue of (2)

$$(17') \quad \gamma^*(x;y) = \gamma(x;y) + R\left\{ e^{2i\phi}\rho^2\left[\phi'(z_0;x)\phi'(z_0;y) - \frac{\phi'(x;y)}{x-z_0} - \frac{\phi'(y;x)}{y-z_0}\right]\right\} + o(\rho^2).$$

Let now C' be a system of smooth curves forming the boundary of an arbitrary partial domain of D which contains x, y and z_0; then the residue theorem yields

$$(24) \quad \frac{1}{2\pi i}\int_{C'} \phi'(z;x)\phi'(z;y)\frac{ds}{z-z_0} = \phi'(z_0;x)\phi'(z_0;y) - \frac{\phi'(x;y)}{x-z_0} - \frac{\phi'(y;x)}{y-z_0},$$

whence, in view of (17′),

$$(25)\qquad \gamma^*(x;y)=\gamma(x;y) + R\left\{ e^{2i\phi}\rho^2\cdot\frac{1}{2\pi i}\int_{C'}\phi'(z;x)\phi'(z;y)\,\frac{ds}{z-z_0}\right\}+o(\rho^2).$$

If the boundary C of D is smooth, we may choose, in particular, $C'=C$; in this case we represent C as in **1**, 1 by means of the length parameter s which insures $|z'(s)|=1$ on every C_ν.

We apply (25) to the functions $\gamma_m(x;y)$ defined by (19); we put

$$(19')\qquad \phi_m(x;y)=\log f_m(x;y),\ \text{i. e.,}\ \gamma_m(x;y)=R\{\phi_m(x;y)\}.$$

By definition of $f_m(x;y)$, $\phi_m(x;y)$ has a constant real part for $z\subset C_\nu$. Hence, along every curve C_ν

$$\frac{d}{ds}\phi_m(z(s);y)=\phi'_m(z;y)\cdot z'$$

is imaginary and its value is, according to Cauchy-Riemann's differential equation,

$$\phi'_m(z;y)z'=-i\frac{\partial}{\partial n_z}\gamma_m(z;y).$$

Hence, we may write (25) in the form

$$(25')\qquad \gamma^*_m(x;y)=\gamma_m(x;y) + \frac{1}{2\pi}\int_C\frac{\partial}{\partial n_z}\gamma_m(z;x)\,\frac{\partial}{\partial n_z}\gamma_m(z;y)R\left\{\frac{i}{z'}\,\frac{e^{2i\phi}\rho^2}{z-z_0}\right\}ds+o(\rho^2).$$

On the other hand, we get obviously for the value of the normal shift by variation (2)

$$(26)\qquad \delta n=R\left\{\frac{i}{z'}\,\frac{e^{2i\phi}\rho^2}{z-z_0}\right\}.$$

Hence, (25′) becomes

$$(27)\qquad \delta\gamma_m(x;y)=\frac{1}{2\pi}\int_C\frac{\partial}{\partial n_z}\gamma_m(z;x)\cdot\frac{\partial}{\partial n_z}\gamma_m(z;y)\,\delta n\,ds,$$

which is just Hadamard's formula for our domain functions. It is derived from (17) for all variations obtainable by superposition of elementary variations (2).

Let $\psi_m(x;y)$ be an analytic function of x, connected with the functions (21) by $\kappa_m(x;y)=R\{\psi_m(x;y)\}$; for $z\subset C_\nu(\nu\neq m)$, $\psi_m(z;y)$ has a constant imaginary part. Hence, on all these curves

$$(28)\qquad \frac{d}{ds}\psi_m(z(s);y)=\psi'_m(z;y)z'=\frac{\partial}{\partial s_z}\kappa_m(z;y).$$

A transformation of (25), analogous to the above, yields by means of (28)

$$(28') \qquad \delta\kappa_m(x;y) = \frac{1}{2\pi}\int_{C_m} \frac{\partial}{\partial n_z}\kappa_m(z;x)\frac{\partial}{\partial n_z}\kappa_m(z;y)\delta n\, ds - \frac{1}{2\pi}\int_{C-C_m}\frac{\partial}{\partial s_z}\kappa_m(z;x)\frac{\partial}{\partial s_z}\kappa_m(z;y)\delta n\, ds.$$

Comparison of (27) and (28′) shows that the same formula (17) appears in its integral representation in very different forms according to the particular boundary conditions satisfied by $\gamma(x;y)$. This shows the advantage of the unifying formula (17).

3. Applications of the variation formula to Green's function.

1. Green's function $g(x;y)$ may be considered as a measure of D with respect to the pair of points x, y in D. In fact, Hadamard's formula (1′) shows the monotonic behavior of $g(x;y)$ as a domain function; for if δn is always positive, i. e., if D is enlarged, $g(x;y)$ is seen to increase, since along the entire boundary we have $\frac{\partial}{\partial n_z} g(z;x) < 0$.

In the neighborhood of y, we have the development

$$(29) \qquad g(x;y) = \log\frac{1}{|x-y|} + \log\frac{1}{d(y)} + O(|x-y|)$$

where $O(\epsilon) \rightarrow 0$ with $\epsilon \rightarrow 0$. If $\infty \subset D$, we have further

$$(29') \qquad g(x;\infty) = \log|x| + \log\frac{1}{d(\infty)} + O\left(\frac{1}{|x|}\right).$$

The functional $d(\infty)$ has been used frequently in the theory of conformal representation as a measure for D or its boundary C. $d(\infty)$ is called transfinite diameter, Robin's constant or capacity constant of C, or of the domain D whose boundary is C (see Fekete, Szegö, Nevanlinna). All domains which may be mapped upon each other by univalent functions $f(x)$, normalized as

$$(30) \qquad f(x) = x + a_0 + \frac{a_1}{x} + \cdots$$

at infinity, have the same measure $d(\infty)$. Analogously, $d(y)$ is a conformal invariant with respect to representations of D which map y on itself and have there the derivative 1. If $f(x)$ has at infinity (or at y) the derivative a, however, D is mapped on a domain with the measure $ad(\infty)$ (or $ad(y)$). In particular, all $d(y)$ behave like lengths with regard to homotheties.

If D is simply connected, it is well known that $d(y)$ is the radius of the circle on the exterior of which D may be mapped by a function with the normalization

$$(30') \qquad f(x) = \frac{1}{x - y} + a_0 + a_1(x - y) + \cdots$$

(or (30) if $y = \infty$).

Introducing into (1′) and (11) the developments (29) and (29′) and letting $x \to y$ we get, by an easy comparison of both sides, the general formulas for the variation of $d(y)$:

$$(31) \qquad \delta \log d(y) = -\frac{1}{2\pi} \int_C \left(\frac{\partial g(z;y)}{\partial n_z} \right)^2 \delta n \, ds$$

$$(31') \qquad \log d^*(y^*) = \log d(y) - R \left\{ e^{2i\phi}\rho^2 \left[p'(z_0;y)^2 - \frac{1}{(y - z_0)^2} \right] \right\} + o(\rho^2).$$

Both formulas remain valid for $y = \infty$.

(31) shows that $d(y)$ is a functional which decreases as the domain increases. If we consider the domain function

$$(32) \qquad \Gamma(x;y) = 2g(x;y) + \log d(x) + \log d(y)$$

we find, by (1′) and (31), its variation formula

$$(32') \qquad \delta\Gamma(x;y) = -\frac{1}{2\pi} \int_C \left[\frac{\partial g(z;x)}{\partial n} - \frac{\partial g(z;y)}{\partial n} \right]^2 \delta n \, ds$$

which proves that $\Gamma(x;y)$ is a decreasing domain function.

2. To demonstrate the applicability of (31′), we deal now with an extremum problem which will become useful later. We choose a boundary continuum C_m of D and a point $y \subset D$; every conformal representation of D of type (30′) transforms C_m into a new continuum $\bar{C}_m$ with a certain transfinite diameter $\delta = \delta(\infty)$. We ask for the extremal values which δ can attain, if all functions (30′), univalent in D, are considered.

To solve this problem of distortion, we start with a domain D with boundary curves $\bar{C}_\nu$, obtained from D by such a univalent function (30′) that δ is maximal. This domain is by no means fixed uniquely; for a representation (30), univalent in the exterior of $\bar{C}_m$, will not change δ and its superposition on any mapping function (30′) will preserve the normalization (30′). This arbitrariness may be avoided by supposing $\bar{C}_m$ to be a circle cen-

tered at the origin. The radius of the circle is necessarily δ, since the transfinite diameter of a circle is equal to its radius. We proceed to determine the shape of the remaining boundary continua $\tilde{C}_\nu$ $(\nu \neq m)$. For this purpose, we choose a continuum $\tilde{C}_l$ $(l \neq m)$, a fixed point z_0 on it and a subcontinuum Γ of $\tilde{C}_l$, containing z_0, of transfinite diameter ρ. All functions (30) which are univalent in the exterior of Γ permit a development

$$(33) \qquad x^+ = \phi(x) = x + k + \frac{a\rho^2}{x - z_0} + \frac{b\rho^3}{(x - z_0)^2} + \cdots$$

valid in every given domain interior to $\tilde{D}$ for Γ sufficiently small, such that $a, b, \cdots$ have bounds independent of Γ. The superposition of a representation (33) on the original mapping of D on $\tilde{D}$ gives a univalent transformation (30') of D on a new domain D^+. This has a boundary continuum C_m^+ arising from $\tilde{C}_m$ by means of (33). Its transfinite diameter $\delta^+ = \delta^+(\infty)$ satisfies, in view of the maximal property of δ, the inequality

$$(34) \qquad \delta^+ \leq \delta.$$

If, on the other hand, $\pi(x; y)$ is an analytic function of x such that $R\{\pi(x; y)\}$ is Green's function for the exterior of $\tilde{C}_m$, we may find δ^+ with the aid of (31'). In fact, the function (33) differs on $\tilde{C}_m$ from

$$(2') \qquad x^* = x + k + \frac{a\rho^2}{x - z_0}$$

only in terms of order $o(\rho^2)$. Hence, in view of (31), applicable in the case $C = \tilde{C}_m$, they cause a variation of $\log \delta$ of this order only, in addition to that caused by (2') and given by (31'). So, we get

$$(35) \qquad \log \delta^+ = \log \delta - R\{a\rho^2\pi'(z_0; \infty)^2\} + o(\rho^2).$$

Since $\tilde{C}_m$ is a circle of radius δ and has as its Green's function $g(x; \infty) = \log \frac{|x|}{\delta}$, we have

$$(36) \qquad \pi(x; \infty) = \log x - \log \delta$$

and, finally, the following formula for the variation of $\log \delta$:

$$(35') \qquad \log \delta^+ = \log \delta - R\left\{ a\rho^2 \cdot \frac{1}{z_0^2} \right\} + o(\rho^2).$$

Comparing (34) with (35') we get the inequality

$$R\left\{a\rho^2\cdot\frac{1}{z_0^{\,2}}\right\}+o(\rho^2)\geq 0, \tag{34'}$$

whatever z_0, Γ and the function (33) may have been.

Now, we have to apply the following lemma (Schiffer 1):

LEMMA. *Let C be a continuum in the x-plane; suppose that there exists an analytic function $s(x)\neq 0$ such that for an arbitrary function (33) univalent in the exterior of an arbitrary subcontinuum Γ of transfinite diameter ρ and containing the arbitrary point z_0 we have*

$$R\{a\rho^2 s(z_0)\}+o(\rho^2)\geq 0. \tag{37}$$

Then C is an analytic curve, expressed by means of a real parameter in the form $x=x(t)$ such that

$$x'(t)^2 s[x(t)]+1=0. \tag{37'}$$

Applying this lemma to (34′) we find $\bar{C}_l$ to be an analytic curve with the differential equation

$$x'(t)^2x(t)^{-2}+1=0 \tag{37''}$$

whence

$$x(t)=k_l e^{it} \qquad k_l=\text{constant}. \tag{37'''}$$

Thus we have proved that δ attains its maximum for the representation of D which transforms C_m into a circle and the other $n-1$ continua C_ν into concentric circular arcs.

As the existence of extremal functions is insured in the case of the problem considered, this theorem proves anew the possibility of this particular type of canonical conformal representation.

Had we raised the question of the minimum of the transfinite diameter δ of all possible $\bar{C}_m$, obtained by representations (30′), our method would have been exactly the same. Only, instead of (34), we would have used the inverse inequality leading to

$$R\left\{a\rho^2\frac{1}{z_0^{\,2}}\right\}+o(\rho^2)\leq 0. \tag{34''}$$

A new application of our lemma shows that in this case also every $\bar{C}_l$ is an analytic curve satisfying now the differential equation

$$x'(t)^2x(t)^{-2}=1 \tag{37 IV}$$

which yields

$$x(t)=\kappa_l e^{t} \qquad \kappa_l=\text{constant}. \tag{37 V}$$

Hence: The minimum of δ is obtained for the representation of D which transforms C_m into a circle and the remaining C_ν into radial segments.

The extremum problem considered above was so easily solved because the function $\pi(x;\infty)$, appearing in the variation formula (35), is the logarithm of a univalent function. Thus, we could apply an auxiliary mapping and give to $\pi(x;\infty)$ a suitable particular form. Most variational problems, concerning the transfinite diameter of a multiply connected domain, are considerably more difficult because the function $p(x;y)$, occurring in (31′), is not connected with univalent functions. In the variational rule of $\gamma_m(x;y)$ and $\kappa_m(x;y)$, however, there appears always the logarithm of a univalent function, which facilitates obviously the treatment of these domain functions.

3. The method of treating extremum problems, just applied, permits an important numerical application in the theory of mapping simply connected domains. Consider all functions (30) which are univalent in $|x| > 1$. They map the unit circle $|x| = 1$ on continua C and our aim is to estimate the distortion of the frontier caused by the mapping. For this purpose, we fix on $|x| = 1$ an arc of length α, say $(e^{-i(\alpha/2)}, 1, e^{i(\alpha/2)})$; let C_α be its corresponding image on C. C_α being a continuum, it is permissible to ask about its transfinite diameter δ. The unit circle has the transfinite diameter 1 and since the latter is preserved by mapping functions (30), C has the same. The transfinite diameter being a decreasing functional of the domain, $C_\alpha \subset C$ yields obviously $\delta \leq 1$. It is easily seen that there exist for every α representations bringing δ arbitrarily near to 1. But there arises the question: What is the minimum of δ? The similarity of this problem to the above is clear; in fact our method of solution will be exactly as before.

Let us suppose that δ attains its minimum for a representation (30) which transforms the unit circle into a continuum $\tilde{C}$ and the arc into the subcontinuum $\tilde{C}_\alpha$. We may suppose $\tilde{C}_\alpha$ to be a circle around the origin with radius δ, since an auxiliary representation (30), univalent in the exterior of C_α, may be superposed on the original mapping of $|x| > 1$ onto $\tilde{D}$ without changing the normalization (30) or the transfinite diameter δ of $\tilde{C}_\alpha$. If z_0 belongs to $\tilde{C}$, but not to $\tilde{C}_\alpha$, we choose a subcontinuum Γ of $\tilde{C}$ of transfinite diameter ρ which contains z_0, but no point of $\tilde{C}_\alpha$. Again we superpose on the original function, mapping $|x| > 1$ on $\tilde{D}$, arbitrary functions (33) which are univalent in the exterior of Γ. We obtain domains D^+ with boundary continuum C^+, containing $C_\alpha{}^+$ as image of the arc. The transfinite diameter δ^+ of $C_\alpha{}^+$ satisfies the inequality $\delta^+ \geq \delta$ in view of the minimum property of δ. Introducing the auxiliary function $\pi(x;y)$, connected with Green's function

of the exterior of $\tilde{C}_a$, we may use (35) for the calculation of $\log \delta^+$. According to our suppposition with resect to $\tilde{C}_a$ we have here too $\pi(x;\infty) = \log x/\delta$. Therefore we find by exactly the same reasoning which led to (37^{V}) that all points of $\tilde{C}$, not belonging to $\tilde{C}_a$, lie on a radius segment in the exterior of $\tilde{C}_a$. By a rotation we may obtain for $\tilde{C}$ finally the following configuration: It consists of the circle $|x| = \delta$ plus the segment $< -\epsilon, -\delta >$, $(\epsilon > \delta)$. The circle corresponds to the arc $(e^{-i(\alpha/2)}, 1, e^{i(\alpha/2)})$ and the segment to the complementary arc.

Thus, the representation is fixed, except for the still unknown values of δ and ϵ. For their determination we consider the representation

$$\xi = x + \delta^2/x \tag{38}$$

of the exterior of $\tilde{C}$ which has the normalization (30). It transforms $\tilde{C}_a$ into the segment $< -2\delta, 2\delta >$ and the rest of $\tilde{C}$ into the segment $< -\epsilon - \delta^2/\epsilon$, $-2\delta >$. On the other hand, the function

$$f(x) = x + 1/x + 2\delta - 2 \tag{38'}$$

is univalent for $|x| > 1$ and is of type (30). It transforms the unit circle into a segment such that the arc $(e^{-i(\alpha/2)}, 1, e^{i(\alpha/2)})$ corresponds to the interval $< 2\delta - 2(1 - \cos\alpha/2), 2\delta >$ on the real axis and the complementary arc to the interval $< 2\delta - 4, 2\delta - 2(1 - \cos\alpha/2) >$. In view of the unicity of the mapping function which transforms the said arcs into *different* contiguous segments of the real axis, we get by comparison

$$2\delta - 2(1 - \cos\alpha/2) = -2\delta, \qquad \epsilon + \delta^2/\epsilon = 4 - 2\delta. \tag{38''}$$

In particular, we get from the first equation (38″)

$$\delta = \sin^2 \alpha/4 \tag{39}$$

for the minimum of the transfinite diameter. Thus, we have proved the following distortion theorem:

Every function (30), *univalent in* $|x| > 1$, *maps an arc of the unit circle with aperture* α *on a continuum with transfinite diameter* $d \geq \sin^2 \alpha/4$.

4. If we start with the domain $|x| > r$, every function (30), univalent in this domain, will map an arc with aperture α on a continuum with transfinite diameter $d \geq r \sin^2 \alpha/4$. This leads to an important application: Let C be an arbitrary continuum with transfinite diameter r; then it may be mapped on the circle $|x| = r$ by a function (30), univalent in its exterior.

Divide C into two continua A and B such that they correspond to two complementary arcs on $|x| = r$ with apertures α and $\beta = 2\pi - \alpha$. Then the foregoing theorem yields:

$$(40) \qquad d(A) \geq r \sin^2 \alpha/4, \qquad d(B) \geq r \sin^2 \beta/4 = r \cos^2 \alpha/4.$$

Since $r = d(C)$ and $C = A + B$, we get from (40) by addition

$$(41) \qquad d(A) + d(B) \geq d(A + B)$$

which establishes in a new way the subadditivity of the transfinite diameter for such a composition (see Schiffer 2).

There are numerous other applications of the theorem of **3** which describes the boundary distortion in case of conformal representation of the unit circle; but we shall not treat them here. We have inserted the proof here only in order to show the use which can be made of (31').

4. The conformal radii $d_m(y)$.

1. It is obvious that the extremal representation (30') of D, transforming C_m into a circle of radius δ and all other C_ν into concentric circular arcs, is closely related to the domain function (19), considered in **2**, **2**. Suppose, indeed, that $f_m(x; y)$ has in the neighborhood of y the development

$$(42) \quad f_m(x; y) = d_m(y)^{-1} \left[\frac{1}{x - y} + a_0 + a_1(x - y) + \cdots \right], \quad d_m(y) > 0;$$

then $d_m(y) \cdot f_m(x; y)$ is of type (30') and defines the extremal representation considered. Hence, $d_m(y)$ is the greatest value for the transfinite diameter which C_m can attain by a representation (30') of D. The area $F(C_m)$ enclosed by the curve C_m may be estimated by means of its transfinite diameter $d(C_m)$ (Pólya 1):

$$(43) \qquad F(C_m) \leq \pi d(C_m)^2.$$

Thus, $\pi d_m(y)^2$ is the greatest area, obtainable for C_m by a representation (30') of D. In virtue of (19) and (42), $d_m(y)$ is connected with $\gamma_m(x; y)$ by

$$(44) \qquad \gamma_m(x; y) = \log \frac{1}{|x - y|} + \log \frac{1}{d_m(y)} + O(|x - y|).$$

If $\infty \subset D$, we have

$$(44') \qquad \gamma_m(x; \infty) = \log |x| + \log \frac{1}{d_m(\infty)} + O\left(\frac{1}{|x|}\right).$$

Thus, $d_m(y)$ is related to $\gamma_m(x; y)$ in exactly the same way as is $d(y)$ tc

Green's function. In view of the above geometric interpretation, we shall call $d_m(y)$ the conformal radius of C_m with respect to C in y. If we are speaking of $d(C_m; C)$, the conformal radius of C_m with respect to C, we shall refer to $d_m(\infty)$. In order that it be defined, we must have $\infty \subset D$; this will be supposed henceforth in this paragraph.

$d(C_m; C)$ measures the continuum C_m, taking into account all the boundary C. It is invariant with respect to representations (30), univalent in D. It is linearly homogeneous with respect to homotheties, as is easily checked. Introducting (44′) into (27), we get by comparison of coefficients

$$(45) \qquad \delta \log d(C_m; C) = -\frac{1}{2\pi} \int_C \left(\frac{\partial}{\partial n} \gamma_m(z; \infty) \right)^2 \delta n\, ds;$$

introducing (44′) into the variation formula (17), with $\phi(x; y) = \log f_m(x; y)$ in our case, we get

$$(46) \qquad \log d(C^*_m; C^*) = \log d(C_m; C) - R\{e^{2i\phi}\rho^2 f''_m(z_0; \infty)^2 f'_m(z_0; \infty)^{-2}\} + o(\rho^2).$$

(45) proves $d(C_m; C)$ to be a decreasing domain function, a fact which follows also easily from its extremal property. Analogously, $d_m(y)$ and even $2\gamma_m(x; y) + \log d_m(x) + \log d_m(y)$ may be shown to be monotonic domain functions; the proof is analogous to that of (32).

2. The following problem requires the application of (46). Divide the boundary C of D into two point sets A and B, each consisting of a finite number of proper continua, and possessing only a finite number of common points. In particular let F be a continuum of A. Consider the conformal radius $d(F; A)$ which is invariant with respect to all representations (30), univalent in the exterior of A. It will change, in general, with respect to mappings (30) of D. We seek the minimum of $d(F; A)$, taking into account all these representations.

In view of the compactness of the family (30) of all functions, univalent in D, there exists at least one function of the family, for which $d(F; A)$ attains its minimum. It maps D, A, F, B on $\tilde{D}$, $\tilde{A}$, $\tilde{F}$, $\tilde{B}$ respectively. It is not uniquely determined since the subsequent superposition of a representation (30), univalent in the exterior of $\tilde{A}$, does not change univalency or normalization of the total mapping function and preserves also $d(\tilde{F}; \tilde{A})$. Thus, it may be supposed that $\tilde{F}$ is a circle around the origin and the other continua of $\tilde{A}$ are concentric circular arcs. In this case, the function $f_F(x; \infty)$ (which corresponds to $f_m(x; \infty)$ in the case $C_m = F$, $C = \tilde{A}$), defining the

canonical representation of $\bar{A}$ with distinction of $\bar{F}$, has the form $xd(\bar{F};\bar{d})^{-1}$, since $d(\bar{F};\bar{A})$ is just the radius of $\bar{F}$. Now, to find the sought for minimum reduces to the determination of the shape of $\bar{B}$.

z_0 being a point of $\bar{B}$ not belonging to $\bar{A}$, let us consider once more the functions (33), univalent in the exterior of a subcontinuum Γ of $\bar{B}$, which contains z_0 but no points of $\bar{A}$. Superposing them on the mapping $D \to \bar{D}$, we get a normalized univalent representation $D \to D^+$. In view of (46) and the simple particular form of $f_F(x;\infty)$, we have, by an argument similar to that which led to (35′),

$$\log d(F^+;A^+) = \log d(\bar{F};\bar{A}) - R\left\{ a\rho^2 \cdot \frac{1}{z_0^2} \right\} + o(\rho^2). \tag{47}$$

Since the minimal property of $d(\bar{F};\bar{A})$ involves

$$\log d(F^+;A^+) \geq \log d(\bar{F};\bar{A}) \tag{48}$$

for an arbitrary choice of the function (33), the lemma of **3**, 2 yields:

$\bar{B}$ consists of segments pointing towards the center of $\bar{F}$.

We have, then, proved the possibility of an interesting type of conformal representation. The boundary C of a domain D may be divided arbitrarily into a finite number of continua, contiguous in a finite number of points only, such that one continuum is mapped on a circle, a given number of continua onto concentric circular arcs and the remainder on radial slits. Every representation of this type solves an extremum problem.

3. The considerations of 2 contain, in particular, the solution of the following problem. Divide the continuum C_1 into the continua A and B which have only two points in common. Let C_A be the aggregate of all points of A and all C_ν $(\nu > 1)$; analogously C_B will consist of B and all C_ν $(\nu > 1)$ Then, we ask for a univalent function (30) in D which imparts to $d(\bar{A};\bar{C}_A)$ its minimum, $\bar{A}$ and $\bar{C}_A$ corresponding to A and C_A by means of the representation. According to 2 we get the solution:

A conformal representation (30) of D which transforms A into a circle $\bar{A}$, B into a contiguous radial segment and the remaining C_ν $(\nu > 1)$ into circular arcs, concentric to the circle, yields the minimum value for $d(\bar{A};\bar{C}_A)$. Obviously, $d(\bar{A};\bar{C}_A)$ is the radius of $\bar{A}$.

This result becomes interesting when compared with the solution of the following question. Consider all functions (30), univalent in D which transform C_1 into a circle. If r denotes the radius of the circle, α the aperture of

the arc corresponding to A, what is the maximum value of the expression $r\sin^2\alpha/4$?

Let $\tilde{D}$ be an image of D yielding the maximum; then, $\tilde{C}_1$ is a circle of radius r, and we may suppose that the arc $(e^{-i(\alpha/2)}, 1, e^{i(\alpha/2)})$ corresponds to A. Again we choose a point z_0 on a fixed $\tilde{C}_\nu$ $(\nu > 1)$ and a function (33), univalent in the exterior of a small subcontinuum of $\tilde{C}_\nu$ containing z_0. In this domain, the function

$$(49) \qquad u = \phi(x) - k + \frac{a\rho^2}{z_0} = x + \frac{a\rho^2 x}{(x - z_0)z_0} + o(\rho^2)$$

is also univalent. This representation deforms $|x| = r$, but for ρ sufficiently small the point $r^2/\bar{z}_0$ remains inside the image of this circumference, i. e., lies in the exterior of the image of D, given by (49). Now, we add the further mapping

$$(49') \qquad u^+ = u - \frac{\bar{a}\rho^2 u^2}{(r^2 - \bar{z}_0 u)\bar{z}_0} = x + \frac{a\rho^2 x}{(x - z_0)z_0} - \frac{\bar{a}\rho^2 x^2}{(r^2 - \bar{z}_0 x)\bar{z}_0} + o(\rho^2)$$

which is univalent in $\tilde{D}$ for ρ small enough. Consider the expression

$$(49'') \qquad \psi(x) = x\left[1 + \frac{a\rho^2}{(x - z_0)z_0} - \frac{\bar{a}\rho^2}{(r^2/x - \bar{z}_0)\bar{z}_0}\right];$$

its modulus for $|x| = r$ is $|\psi(x)| = r + o(\rho^2)$. Hence, adding an additional correction term of order $o(\rho^2)$, we infer that $u^+(x)$, so corrected, transforms the circle $|x| = r$ into itself. At the same time, the points $re^{i(\alpha/2)}$ and $re^{-i(\alpha/2)}$ on it are mapped on the points

$$(50) \qquad re^{i(\alpha_{1/2})} = re^{\pm i(\alpha/2)}\left[1 + 2iI\left\{\frac{a\rho^2}{(re^{\pm i(\alpha/2)} - z_0)z_0}\right\}\right] + o(\rho^2).$$

Thus, the new arc, corresponding to A, has the aperture

$$(51) \qquad \alpha^+ = \alpha_1 - \alpha_2 = \alpha - 4r\sin\alpha/2 \times R\left\{\frac{a\rho^2}{z_0(r^2 + z_0{}^2 - 2rz_0\cos\alpha/2)}\right\} + o(\rho^2).$$

The function (49′), though univalent in $\tilde{D}$, has not yet the normalization (30), but must be divided by $\left(1 + \dfrac{\bar{a}\rho^2}{\bar{z}_0{}^2} + o(\rho^2)\right)$ for this purpose. This operation preserves the aperture α^+, but yields the radius

$$(52) \qquad r^+ = r\left[1 - R\left\{\frac{a\rho^2}{z_0{}^2}\right\}\right] + o(\rho^2).$$

The expression $r^{+}\sin^2\alpha^{+}/4$ belongs to a normalized function, univalent in $\bar{D}$. By (51) and (52) we have

$$(53)\qquad r^{+}\sin^2\alpha^{+}/4 = r\sin^2\alpha/4\left[1 - R\left\{a\rho^2\frac{(r+z_0)^2}{z_0^2(r^2+z_0^2-2rz_0\cos\alpha/2)}\right\}\right] + o(\rho^2).$$

Since $\bar{D}$ was supposed to be an extremal domain, we have for every choice of (33)

$$(54)\qquad R\left\{a\rho^2\frac{(r+z_0)^2}{z_0^2(r^2+z_0^2-2rz_0\cos\alpha/2)}\right\} + o(\rho^2) \geq 0.$$

Applying now the lemma of **3**, 2, we find every $\bar{C}_\nu$ $(\nu > 1)$ to be an analytic curve $x = x(t)$, satisfying the differential equation

$$(55)\qquad \frac{x'(t)^2}{x(t)^2}\,\frac{[r+x(t)]^2}{[re^{i(\alpha/2)}-x(t)][re^{-i(\alpha/2)}-x(t)]} + 1 = 0.$$

Thus, the extremal domain $\bar{D}$ is bounded by a circle $\bar{C}_1$ of radius r and by $n-1$ curves (55). In order to understand better the structure of $\bar{D}$, consider the function $\zeta(x)$, univalent in the exterior of $\bar{C}_1$ and of type (30), which maps the arc $\bar{A}$ of $\bar{C}_1$, corresponding to A, on a circle around the origin, while its complement $\bar{B}$ becomes a segment of the negative axis. By these requirements, $\zeta(x)$ is fixed and may be calculated in an elementary way The radius of the new circle is $r\sin^2\alpha/4$. The expression $x^2\zeta'(x)^2\zeta(x)^{-2}$ is a regular function for $|x| > r$; it is positive on $\bar{A}$, negative on $\bar{B}$, has in $re^{i(\alpha/2)}$ and $re^{-i(\alpha/2)}$ simple poles and in $-r$ a double zero point, as is easily seen from geometric considerations. Taking into account, further, the normalization at infinity, we get by means of Schwarz's principle of reflection

$$(56)\qquad \frac{x^2\zeta'(x)^2}{\zeta(x)^2} = \frac{(x+r)^2}{(x-re^{i(\alpha/2)})(x-re^{-i(\alpha/2)})}.$$

With its aid, (55) may be written in simple form, if we put $\zeta[x(t)] = \xi(t)$:

$$(55')\qquad \frac{\xi'(t)^2}{\xi(t)^2} + 1 = 0;\ \text{i.e.,}\ \xi(t) = \rho_\nu e^{it}, \qquad 0 < \rho_\nu = \text{constant}.$$

The final result may be formulated in the following way:

Among all univalent representations (30) of D which transform A into a circle $\bar{A}$ and B into a contiguous radial segment, the maximal radius for $\bar{A}$ is attained, if all other C_ν $(\nu > 1)$ become circular arcs, concentric to $\bar{A}$.

Thus, the same representation imparts to $d(A; C_A)$ its minimum and to $r\sin^2\alpha/4$ its maximum which has the value $d(A; C_A)$. Hence, we get the inequality

(57) $$d(A; C_A) \geq r \sin^2 \alpha/4$$

valid for every representation (30) which transforms C_1 into a circle. By means of such representations, there corresponds to A an arc of aperture α, and to its complement B on C_1 an arc of aperture $\beta = 2\pi - \alpha$. In view of (57) and the analogous formula for B

(57′) $$d(B; C_B) \geq r \sin^2 \beta/4 = r \cos^2 \alpha/4,$$

we get by addition

(58) $$d(A; C_A) + d(B; C_B) \geq r.$$

Since the maximum of r is given by the conformal radius $d(C_1; C)$ (4, 1), we get

(58′) $$d(A; C_A) + d(B; C_B) \geq d(A + B; C)$$

in complete analogy to the subadditivity formula (41) for the transfinite diameter.

5. The construction of $\gamma_m(x; y)$ by means of Green's function.

1. The functions $\gamma_m(x; y)$ may be constructed explicitly by means of Green's function $g(x; y)$ and other functions derived from it. In this way, we may derive the variation formula (17) for $\gamma_m(x; y)$ directly from (11). For this construction, we have to study the function $p(x; y)$, introduced in 1, 2 by (7), more thoroughly. Though its real part $g(x; y)$ is uniform in D, it has with respect to circuits around C_ν the periods

(59) $$2\pi i \omega_\nu(y) = i \int_{C_\nu} \frac{\partial}{\partial n_z} g(y; z)\, ds_z.$$

$\omega_\nu(y)$ is a real-valued harmonic function for $y \subset D$; by virtue of its integral representation it attains on C_ν the value 1, on the remaining C_μ the value 0. $\omega_\nu(y)$ is called the harmonic measure of C_ν in y with respect to D (Nevanlinna 1). Obviously, $\omega_\nu(y)$ is the real part of the analytic function

(60) $$w_\nu(y) = \frac{1}{2\pi} \int_{C_\nu} \frac{\partial}{\partial n_z} p(y; z)\, ds_z.$$

$w_\nu(y)$ is regular in D and possesses periods with respect to circuits around every C_μ; they are

(61) $$2\pi i P_{\mu\nu} = i \int_{C_\mu} \frac{\partial \omega_\nu(y)}{\partial n_y}\, ds_y = \frac{i}{2\pi} \int_{C_\mu} \int_{C_\nu} \frac{\partial^2 g(y; z)}{\partial n_y \partial n_z}\, ds_y ds_z = 2\pi i P_{\nu\mu}.$$

The matrix $(P_{\mu\nu})$ is thus seen to be symmetric.

Consider, for $x \subset D$, the harmonic function

(62) $$v(x) = \sum_{\nu=1}^{n} c_\nu \omega_\nu(x)$$

which has on the curve C_ν the boundary value c_ν. By Green's theorem we have

(63) $$\frac{1}{2\pi} \int\int_D (\operatorname{grad} v)^2 \, d\tau = -\frac{1}{2\pi} \int_C v \frac{\partial v}{\partial n} \, ds = -\sum_{\mu,\nu=1}^{n} P_{\mu\nu} c_\mu c_\nu.$$

This expression is non-negative and vanishes only for $\operatorname{grad} v \equiv 0$, i. e., $v(x) = \text{const}$. In this case, all c_ν are necessarily equal, which leads to the theorem:

The quadratic form $\sum_{\mu,\nu=1}^{n} P_{\mu\nu} c_\mu c_\nu$ *is non-positive and vanishes only if all* c_ν *are equal.*

The harmonic function $\sum_{\nu=1}^{n} \omega_\nu(x)$ has everywhere on the boundary of D the value 1 and coincides, therefore, in D with the constant 1. Hence, the analytic function $\sum_{\nu=1}^{n} w_\nu(x)$ is also constant and its period with respect to a circuit around each C_μ vanishes:

(64) $$2\pi i \sum_{\nu=1}^{n} P_{\mu\nu} = 0 \qquad (\mu = 1, \cdots, n).$$

Thus, the sum over every row (or column) in the matrix $(P_{\mu\nu})$ is zero and, consequently, the determinant of this matrix vanishes.

Striking out the n-th row and the n-th column of the matrix $(P_{\mu\nu})$ we obtain a matrix with a corresponding negative-definite form, thus with non-vanishing determinant. Let its inverse matrix be $(p_{\mu\nu}^{(n)})$, the upper index indicating that μ and ν do not assume the value n.

2. By means of the previously defined concepts we may easily construct the function $\gamma_n(x; y)$. We consider the analytic function of $x \subset D$, for $y \subset D$ fixed,

(65) $$\phi_n(x; y) = p(x; y) - \sum_{\mu,\nu=1}^{n-1} p_{\mu\nu}^{(n)} w_\mu(x) \omega_\nu(y).$$

It has by (59) and (61) the period

(65′) $$\pi_\sigma = 2\pi i [\omega_\sigma(y) - \sum_{\mu,\nu=1}^{n-1} p_{\mu\nu}^{(n)} P_{\mu\sigma} \omega_\nu(y)]$$

with respect to a circuit around C_σ. For $\sigma \neq n$, the definition of $(p_{\mu\nu}^{(n)}$

insures $\pi_\sigma = 0$. In order to calculate π_n we apply (64) and express $P_{\mu\nu}$ by means of all $P_{\mu\tau}$ $(\tau \neq n)$. It is then easily seen that

$$\pi_n = 2\pi i[\omega_n(y) + \sum_{\tau=1}^{n-1} \omega_\tau(y)] = 2\pi i. \tag{65''}$$

The function $\phi_n(x;y)$, regular for $x \subset D$ with exception of the logarithmic pole at y, does not change, therefore, for circuits around C_ν $(\nu \neq n)$ and increases by $2\pi i$ for a circuit around C_n. Its real part is constant on every C_ν; in particular, it vanishes on C_n. Hence, $\phi_n(x;y) = \log f_n(x;y)$, $f_n(x;y)$ mapping D onto the exterior of the unit circle, slit along concentric circular arcs. Thus, in view of (19), (59), (60) and (65)

$$\gamma_n(x;y) = R\{\phi_n(x;y)\} = g(x;y) - \sum_{\mu,\nu=1}^{n-1} p_{\mu\nu}{}^{(n)}\omega_\mu(x)\omega_\nu(y). \tag{66}$$

Analogously, we may construct all the functions $\gamma_m(x;y)$ $(1 \leq m \leq n-1)$, by defining the matrix $(p_{\mu\nu}{}^{(m)})$, inverse to the matrix obtained by striking out in $(P_{\mu\nu})$ the m-th row and the m-th column.

3. If y and z are two fixed points in D, consider the analytic function of x

$$\phi_n(x;y,z) = \phi_n(x;y) - \phi_n(x;z). \tag{67}$$

It is uniform for every circuit, since the periods for a circuit around C_n cancel each other exactly. It has logarithmic poles in y and z and a constant real part on every boundary continuum C_ν, vanishing in particular on C_n. Thus, it is the logarithm of the function which maps D on a plane, slit along concentric circular arcs, so that y goes to infinity, z to zero and that C_n corresponds to an arc of the unit circle.

This representation is connected with the following extremum problem: Determine a function (30′), univalent in D, which yields the maximum value for $|f'(z)|$, $z \subset D$ (see de Possel, Rengel).

Let $\tilde{D}$ be the domain belonging to the extremal function, $\tilde{C}_\nu$ its boundary curves. Using a point z_0 on $\tilde{C}_\nu$, we consider the functions (33), univalent in the exterior of a subcontinuum Γ of $\tilde{C}_\nu$ with transfinite diameter ρ which contains z_0. Superposing such a function on the extremal representation $D \to \tilde{D}$, gives a resultant conformal representation of type (30′) of D, since

$$\phi[f(x)] = f^*(x) = f(x) + \frac{a\rho^2}{f(x) - z_0} + o(\rho^2) \tag{68}$$

has at $x = y$ the principal part $\dfrac{1}{x-y}$. Further, we have

$$(69) \qquad f^{*\prime}(z) = f'(z)\left[1 - \frac{a\rho^2}{(f(z) - z_0)^2} + o(\rho^2)\right].$$

In view of the maximum property of $|f'(z)|$ there holds for every function (38)

$$(70) \qquad R\left\{\frac{a\rho^2}{(f(z) - z_0)^2}\right\} + o(\rho^2) \leq 0.$$

Hence, in virtue of Lemma **3**, 2, every C_ν is an analytic curve $x = x(t)$ with the differential equation

$$(71) \qquad x'(t)^2[f(z) - x(t)]^{-2} + 1 = 0.$$

Without restriction of generality we may suppose $f(z) = 0$, since addition of a constant to $f(x)$ changes neither its derivative nor its normalization (30′). Thus, all $\tilde{C}_\nu$ have the form

$$(70') \qquad x(t) = c_\nu e^{it}, \qquad 0 < c_\nu = \text{constant},$$

i. e., they are all circular arcs around the common center $f(z) = 0$.

Thus, the extremal function $f(x)$, yielding this representation, coincides, by the unicity theorem for this type of conformal representation, with a constant multiple of $\exp\{\phi_n(x; y, z)\}$. Hence,

$$(72) \qquad \log|f(x)| = k + g(x; y) - g(x; z) - \sum_{\mu,\nu=1}^{n-1} p_{\mu\nu}{}^{(n)}\omega_\mu(x)[\omega_\nu(y) - \omega_\nu(z)].$$

In order to eliminate the constant k, we combine (72) with (29) and (30′); for $x \to y$, we get

$$(72') \qquad 0 = k + \log\frac{1}{d(y)} - g(y; z) - \sum_{\mu,\nu=1}^{n-1} p_{\mu\nu}{}^{(n)}\omega_\mu(y)[\omega_\nu(y) - \omega_\nu(z)].$$

On the other hand, for $x \to z$, we get in the same way by means of the development $f(x) = f'(z)(x - z) + \cdots$, valid in the neighborhood of the point z,

$$(72'') \qquad \log|f'(z)| = k + g(z; y) - \log\frac{1}{d(z)} - \sum_{\mu,\nu=1}^{n-1} p_{\mu\nu}{}^{(n)}\omega_\mu(z)[\omega_\nu(y) - \omega_\nu(z)].$$

By subtracting (72′) from (72″) we get

$$(73) \qquad \log|f'(z)| = 2g(z; y) + \log d(z) + \log d(y) + \sum_{\mu,\nu=1}^{n-1} p_{\mu\nu}{}^{(n)}[\omega_\mu(y) - \omega_\mu(z)][\omega_\nu(y) - \omega_\nu(z)].$$

The last term is non-positive, since $(p_{\mu\nu}^{(n)})$ is the inverse of a matrix with a non-positive quadratic form. The maximum for $|f'(z)|$ is $\geq \dfrac{1}{|z-y|^2}$, since this value is always attained in the particular case $f(x) = \dfrac{1}{x-y} - \dfrac{1}{z-y}$. Thus we have proved for the function (32) the interesting inequality

$$\Gamma(x;y) = 2g(x;y) + \log d(x) + \log d(y) \geq \log \frac{1}{|x-y|^2}. \tag{74}$$

From (66) we obtain, by comparing coefficients

$$\log \frac{1}{d_n(y)} = \log \frac{1}{d(y)} - \sum_{\mu,\nu=1}^{n-1} p_{\mu\nu}^{(n)} \omega_\mu(y) \omega_\nu(y). \tag{75}$$

Applying this identity and (66) to (73), we get

$$\log |f'(z)| = 2\gamma_n(z;y) + \log d_n(z) + \log d_n(y). \tag{76}$$

Thus, we have established a simple relation between the domain function $\gamma_n(z;y)$ and the maximal derivative of all functions (30′), univalent in D. It makes obvious the fact, established in **4**, 1, that the right hand side of (76) is a monotone decreasing function of the domain (diminution of the family of competing functions decreases the maximum).

4. After having shown the significance of the expressions $w_\nu(x)$ and $P_{\mu\nu}$, we ask now about their variational formulae. We shall derive them from (11) and get in this way anew the variation formula for $\gamma_n(x;y)$. The variation (2) transforms the domain D with boundary curves C_ν and Green's function $g(x;y)$ into a domain D^* with boundary curves C^*_ν and Green's function $g^*(x;y)$. The connection between Green's functions of both domains is established by (11). If the point x describes a circuit around C_ν, its image point x^* turns around the corresponding C^*_ν, and, by (11) and (59), the period of $p^*(x^*;y^*)$ with respect to this circuit is

$$\begin{aligned}(77)\quad 2\pi i\omega^*_\nu(y^*) &= i\int_{C_\nu} \frac{\partial}{\partial n_x} g^*(y^*;x^*)\,ds_x \\ &= i\int_{C_\nu} \frac{\partial g(y;x)}{\partial n_x}\,ds_x + iR\{e^{2i\phi}\rho^2 p'(z_0;y)\int_{C_\nu} \frac{\partial p'(z_0;x)}{\partial n_x}\,ds_x\} + o(\rho^2)\end{aligned}$$

whence, by virtue of (59) and (60),

$$\omega^*_\nu(y^*) = \omega_\nu(y) + R\{e^{2i\phi}\rho^2 p'(z_0;y) w'_\nu(z_0)\} + o(\rho^2). \tag{78}$$

This formula yields the variation of the harmonic measure $\omega_\nu(y)$. Using (61),

we derive herefrom the variation formula for the $P_{\mu\nu}$. In fact, we have from (78):

$$(79) \qquad P^*_{\mu\nu} = \frac{1}{2\pi} \int_{C_\mu} \frac{\partial}{\partial n_y} \omega^*_\nu(y^*) \, ds_y$$
$$= P_{\mu\nu} + R\{e^{2i\phi}\rho^2 w'_\nu(z_0) \cdot \frac{1}{2\pi} \int_{C_\mu} \frac{\partial p'(z_0; y)}{\partial n_y} ds_y\} + o(\rho^2),$$

whence, in view of (60),

$$(80) \qquad P^*_{\mu\nu} = P_{\mu\nu} + R\{e^{2i\phi}\rho^2 w'_\mu(z_0) w'_\nu(z_0)\} + o(\rho^2).$$

The expressions $p_{\mu\nu}^{(n)}$ are defined by the system of equations

$$(81) \qquad \sum_{\lambda=1}^{n-1} p_{\mu\lambda}^{(n)} P_{\lambda\sigma} = \delta_{\mu\sigma} = \begin{cases} 1 & \mu = \sigma \\ 0 & \mu \neq \sigma \end{cases} \qquad (\mu, \sigma = 1, \cdots, n-1).$$

We derive from (80) and (81) the following system of equations for $p_{\mu\nu}^{(n)*}$:

$$(82) \qquad \delta_{\mu\sigma} = \sum_{\lambda=1}^{n-1} p_{\mu\lambda}^{(n)*} P^*_{\lambda\sigma} = \sum_{\lambda=1}^{n-1} p_{\mu\lambda}^{(n)*} P_{\lambda\sigma}$$
$$+ R\{e^{2i\phi}\rho^2 w'_\sigma(z_0) \sum_{\lambda=1}^{n-1} p_{\mu\lambda}^{(n)} w'_\lambda(z_0)\} + o(\rho^2).$$

Multiplying both sides with $p_{\sigma\nu}^{(n)}$ and summing up with respect to σ, we get from (81), in view of the symmetry property $p_{\mu\nu}^{(n)} = p_{\nu\mu}^{(n)}$

$$(83) \quad p_{\mu\nu}^{(n)*} + R\{e^{i2\phi}\rho^2 \sum_{\sigma=1}^{n-1} p_{\nu\sigma}^{(n)} w'_\sigma(z_0) \cdot \sum_{\lambda=1}^{n-1} p_{\mu\lambda}^{(n)} w'_\lambda(z_0)\} + o(\rho^2) = p_{\mu\nu}^{(n)},$$

i. e.,

$$(83') \quad p_{\mu\nu}^{(n)*} = p_{\nu\mu}^{(n)} - R\{e^{2i\phi}\rho^2 \sum_{\kappa=1}^{n-1} p_{\mu\kappa}^{(n)} w'_\kappa(z_0) \cdot \sum_{\lambda=1}^{n-1} p_{\nu\lambda}^{(n)} w'_\lambda(z_0)\} + o(\rho^2).$$

From (66), (11), (78) and (83′) we get finally

$$(84) \quad \gamma^*_n(x^*; y^*) = \gamma_n(x; y)$$
$$+ R\{e^{2i\phi}\rho^2[p'(z_0; x) - \sum_{\mu,\nu=1}^{n-1} p_{\mu\nu}^{(n)} w'_\mu(z_0)\omega_\nu(x)][p'(z_0; y)$$
$$- \sum_{\mu,\nu=1}^{n-1} p_{\mu\nu}^{(n)} w'_\mu(z_0)\omega_\nu(y)] + o(\rho^2),$$

and using definition (65)

$$(84') \quad \gamma^*_n(x^*; y^*) = \gamma_n(x; y) + R\{e^{2i\phi}\rho^2\phi'_n(z_0; x)\phi'_n(z_0; y)\} + o(\rho^2).$$

This coincides exactly with the variation formula (17) for $\gamma_n(x; y)$.

6. The functionals of a Riemann surface and their variation.

1. The above formal relations between the functionals of a domain D: $g(x;y)$, $\omega_\nu(y)$ and $P_{\mu\nu}$ recall similar connections between the elementary integrals on a Riemann surface and their periods. These analogies are not accidental; they result from the close relations between the theory of conformal representation of multiply connected domains and the theory of Riemann surfaces (see Schottky).

The elementary integrals are functionals of the Riemann surface and we want to study how their variation depends upon that of the surface. To this purpose, we have first to recall certain concepts from the theory of Riemann surfaces.

If the Riemann surface P has the genus $p \geqq 1$, there exists on it a canonical system of $2p$ closed rectifiable curves a_j, b_j $(j = 1, \cdots, p)$, such that every pair a_j, b_j has exactly one point of intersection, while no two curves of this system have further common points. Let $v_j(x)$ be analytic on P (i. e., it has a convergent development into power series at every point of P), with the period 1 with respect to a circuit on a_j and the period 0 with respect to circuits on every other a_k. $v_j(x)$ is fixed by this condition, up to an additive constant and is called the j-th elementary integral of first kind on P. The periods of all $v_j(x)$ with respect to circuits on the curves b_k form the matrix (π_{jk}).

Let $t(x;y)$ be an analytic function of x on P, the point y excepted where it has a simple pole with residue 1. Let $t(x;y)$ remain unchanged if x describes any curve a_j. These conditions fix $t(x;y)$ up to an additive constant; $t(x;y)$ is called an elementary integral of the second kind on P.

Let $w(x;y,z)$ be an analytic function of x *on* P, the points y and z excepted where it has logarithmic poles with the principal part $\log \dfrac{x-z}{x-y}$. Let $w(x;y,z)$ not change if x describes a curve a_j. These conditions fix $w(x;y,z)$ up to an additive constant. $w(x;y,z)$ is called an elementary integral of the third kind on P.

The elementary integrals of the first and second kind are uniform functions of x on the surface $\tilde{\mathrm{P}}$, obtained from P by cutting it along all the curves of the system a_j, b_j. The boundary Σ of P consists of the curves a_j, b_j; but every side of each curve being counted separately, every curve appears twice in Σ. Let $f(x)$ be uniform and meromorphic on $\tilde{\mathrm{P}}$; then

$$S\{f(x)\} = \frac{1}{2\pi i}\int_\Sigma f(x)\,dx \tag{85}$$

represents the sum of all residues of $f(x)$ on $\bar{\mathrm{P}}$. If, in particular, $f(x)$ is a linear aggregate of a finite number of elementary integrals with coefficients uniform on P, we may, on the other hand, evaluate the expression (85) by using the characteristic periodicity relations of the elementary integrals, mentioned above. Riemann's method of boundary integration compares both results and obtains in this way relations between elementary integrals and their periods.

Since we shall use the result later on, we apply Riemann's method to

$$(86) \qquad S\{t(x;u)w'(x;y,z)\} = \frac{1}{2\pi i}\int_{\Sigma} t(x;u)w'(x;y,z)\,dx, \qquad u \subset \tilde{\mathrm{P}}$$

(the dash denotes as usual the derivative with respect to the first variable). Integrating along Σ, we have to pass every curve a_j and b_j twice, but in opposite directions. We get from one bank of the curve a_j to the other by following the curve b_j, and from one bank of b_j to the other by describing a_j (starting always in their common point). $t(x;u)$ having the period zero with respect to every circuit a_j, it has the same value on both banks of every b_j; hence, the integrations along the two banks of b_j in (86) cancel each other. The values of $t(x;u)$ on both banks of a_j differ by a constant period, depending, however, upon u, say $p_j(u)$. Thus

$$(86') \qquad S\{t(x;u)w'(x;y,z)\} = -\sum_{k=1}^{p}\frac{1}{2\pi i}\int_{a_k} w'(x;y,z)\,dx\cdot p_k(u).$$

Now, $w(x;y,z)$ does not change after describing a full circuit a_k; hence, each integral in (86′) vanishes, and their sum also.

On the other hand, we have, by virtue of the residue theorem,

$$(87) \qquad S\{t(x;u)w'(x;y,z)\} = w'(u;y,z) - [t(y;u) - t(z;u)].$$

This being zero by the foregoing argument, we have established the following relation between elementary integrals of the second and the third kind:

$$(88) \qquad w'(x;y,z) = t(y;x) - t(z;x).$$

In the same way the following relations are proved: The equalities

$$(89) \qquad \pi_{jk} = \pi_{kj};$$

the periodicity relations

$$(90) \qquad \frac{1}{2\pi i}\int_{b_j} t'(x;y)\,dx = -v'_j(y),$$

$$(91) \qquad \frac{1}{2\pi i}\int_{b_j} w'(x;y,z)\,dx = -[v_j(y) - v_j(z)],$$

and the symmetry relations

$$t'(x;y) = t'(y;x) \tag{92}$$

$$w(x;y,z) - w(u;y,z) = w(y;x,u) - w(z;x,u). \tag{93}$$

2. After these preparations let us examine the change of the elementary integrals due to a variation of P. It is known that they depend continuously upon P (see Ritter, Koebe 2). But no explicit formula, describing the dependency, has been given as yet. We consider the following particular variations of P.

Let $f(x)$ be uniform and meromorphic on P with simple poles z_ν $(\nu = 1, \cdots, N)$ and residues r_ν at z_ν. For sake of simplicity we suppose that no z_ν coincides with a branch point of P or lies on a curve of our canonical system. Consider the function

$$x^* = x + \rho^2 f(x). \tag{94}$$

Fix around each z_ν a small circumference k_ν, not containing any branch point of P or a point of the canonical curve system, such that no two circles overlap. Discarding the interiors of k_ν from P, we obtain a surface P_0 with N holes. If ρ is sufficiently small, P_0 is mapped by (94) in a one-to-one manner on a surface P^*_0 with N holes, which can be completed to a closed Riemann surface P^* (see Schiffer 3).

Our purpose is to calculate for this variation $P \rightarrow P^*$ the corresponding variation formulae for the elementary integrals and their periods. We remark that P and P^* are of the same genus and that the variation (94) transforms the system a_j, b_j again into a canonical system a^*_j, b^*_j on P^*. The functions $v^*_j(x)$, $t^*(x;y)$, $w^*(x;y,z)$ being the elementary integrals on P^*, the functions $v^*_j(x^*(x))$, $t^*(x^*(x);y^*(y))$ and $w^*(x^*(x);y^*(y),z^*(z))$ are given on P_0 by means of (94). If x describes the system a_j, b_j, its image x^* does the same with respect to a^*_j, b^*_j. We establish now the equality

$$\frac{1}{2\pi i} \int_\Sigma t(x;u) \frac{d}{dx} w^*(x^*;y^*,z^*)\,dx = \sum_{j=1}^{p} v'_j(u) \int_{a_j} \frac{d}{dx} w^*(x^*;y^*,z^*)\,dx, \tag{95}$$

by means of the periodicity properties of $t(x;u)$. In fact, $t(x;u)$ has the periods zero with respect to the circuits a_j and, by (90), the periods $-2\pi i\, v'_j(u)$ regarding the b_j. Further, $w^*(x^*;j^*,z^*)$ does not change, if x^* makes a full circuit around a^*_j, i. e., if x describes a_j. Thus,

$$\frac{1}{2\pi i} \int_\Sigma t(x;u) \frac{d}{dx} w^*(x^*;y^*,z^*)\,dx = 0. \tag{96}$$

If, on the other hand, u, y, z lie in P_0, we may evaluate the last integral by Cauchy's theorem and so derive from (96), the equality

$$(97) \qquad 0 = \frac{d}{du} w^*(u^*; y^*, z^*) - [t(y; u) - t(z; u)] + \sum_{\nu=1}^{N} \frac{1}{2\pi i} \int_{k\nu} t(x; u) \frac{d}{dx} w^*(x^*; y^*, z^*) \, dx.$$

By elementary development into series, using (94) and the residue theorem, we find

$$(98) \quad \frac{1}{2\pi i} \int_{k\nu} t(x; u) \frac{d}{dx} w^*(x^*; y^*, z^*) \, dx = - r_\nu \rho^2 t'(z_\nu; u) w'(z_\nu; y, z) + o(\rho^2).$$

In view of (88) and (98) we may write, instead of (97),

$$(99) \quad \frac{d}{du} w^*(u^*; y^*, z^*) = \frac{d}{du} w(u; y, z) + \sum_{\nu=1}^{N} r_\nu \rho^2 t'(z_\nu; u) w'(z_\nu; y, z) + o(\rho^2).$$

By virtue of (92), (99) may be integrated with respect to u from x_0 to x. Applying once more (88), we find

$$(100) \qquad w^*(x^*; y^*, z^*) - w^*(x_0^*; y^*, z^*) = w(x; y, z) - w(x_0; y, z) + \sum_{\nu=1}^{N} r_\nu \rho^2 w'(z_\nu; x, x_0) w'(z_\nu; y, z) + o(\rho^2).$$

Now, let x describe the curve b_j; then x^* will describe b^*_j and, in view of (91), (88) and (90), we get by comparison of the periods on both sides of (100):

$$(101) \qquad v^*_j(y^*) - v^*_j(z^*) = v_j(y) - v_j(z) + \sum_{\nu=1}^{N} r_\nu \rho^2 v'_j(z_\nu) w'(z_\nu; y, z) + o(\rho^2).$$

Next, let y describe the curve b_k and compare once more the increments on both sides:

$$(102) \qquad \pi^*_{jk} = \pi_{jk} - \sum_{\nu=1}^{N} r_\nu \rho^2 v'_j(z_\nu) v'_k(z_\nu) + o(\rho^2).$$

The great similarity between (100), (101) and (102), on the one hand, and (11), (72) and (80), on the other, is obvious. The above formulas determine explicitly the variation of the elementary integrals and their periods on Riemann surfaces. To their applications on extremum problems concerning Riemann surfaces we hope to return elsewhere.

Einstein Institute,
The Hebrew University,
Jerusalem, Palestine.

BIBLIOGRAPHY.

S. Bergman. (1) *Partial differential equations, Advanced topics* (*Conformal mapping of multiply connected domains*), Publication of Brown University, Providence, R. I., 1941.

———. (2) A remark on the mapping of multiply connected domains, *American Journal of Mathematics*, vol. 68 (1946), pp. 20-28.

M. Biernacki. "Sur quelques majorantes de la théorie des fonctions univalentes, Comptes Rendus, vol. 201 (1935), pp. 256-258.

M. Fekete. "Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten," *Math. Zeitschrift*, vol. 17 (1923), pp. 228-249.

H. Grötzsch. "Über die Verzerrung bei schlichter konformer Abbildung mehrfach-zusammenhängender schlichter Bereiche II," *Berichte Leipzig*, vol. 81 (1929), pp. 217-221.

J. Hadamard. "Mémoire sur le problème d'analyse relatif à l'équilibre des plaques élastiques encastrées," *Mémoires présentés par divers savants à l'Académie des Sciences*, vol. 33 (1908).

G. Julia. "Sur une équation aux dérivées fonctionelles liée à la représentation conforme," *Annales de l'Ecole Normale* (3), vol. 39 (1922), pp. 1-28.

P. Koebe. (1) Abhandlungen zur Theorie der konformen Abbildung IV," *Acta Math.*, vol. 41 (1918), pp. 305-344.

———. (2) "Über die Uniformisierung algebraischer Kurven IV," *Math. Annalen*, vol. 75 (1914), pp. 42-129.

P. Lévy. *Leçons d'Analyse fonctionelle*, Paris, 1922.

R. Nevanlinna. *Eindeutige analytische Funktionen*, Berlin, 1936.

G. Pólya. "Beitrag zur Verallgemeinerung des Verzerrungssatzes auf mehrfach-zusammenhängende Gebiete II," *Sitzungsberichte Akad. Berlin* (1928), pp. 280-282.

R. de Possel. "Sur quelques problèmes de représentation conforme," *Comptes Rendus*, vol. 194 (1932), pp. 42-44.

E. Rengel. "Existenzbeweise für schlichte Abbildungen mehrfach zusammenhängender Bereiche auf gewisse Normalbereiche," *Jahresb. D. Math. Ver.*, vol. 44 (1934), pp. 51-55.

E. Ritter. "Die Stetigkeit der automorphen Funktionen bei stetiger Abänderung des Fundamentalbereiches," *Math. Annalen*, vol. 45 (1894), pp. 473-544 and vol. 46 (1895), pp. 200-248.

M. Schiffer. "A method of variation within the family of simple functions," *Proceedings of the London Mathematical Society* (2), vol. 44 (1938), pp. 432-449.

———. (2) "On the subadditivity of the transfinite diameter," *Proceedings of the Cambridge Philosophical Society*, vol. 37 (1941), pp. 373-383.

———. (3) "Variation of the Green function and theory of the p-valued functions," *American Journal of Mathematics*, vol. 65 (1943), pp. 341-360.

———. (4) "On the kernal function of a system of ortho-normal functions," to appear later.

E. Schottky. "Über die konforme Abbildung mehrfach zusammenhängender ebener Flächen," *Journ. f. reine u. angew. Math.*, vol. 83 (1877), pp. 300-351.

G. Szegö. "Bemerkung zu einer Arbeit von Herrn M. Fekete," *Math. Zeitschr.*, vol. 21 (1924), pp. 203-208.

V. Volterra. *Theory of Functionals*, London, 1931.

EULER-KNOPP SUMMABILITY OF CLASSES OF CONVERGENT SERIES.*

By M. S. Macphail.

In connection with the study of Euler-Knopp methods of summability recently made by R. P. Agnew [1], the theorem of the present note may be of interest. A given series $u_0 + u_1 + \cdots$ with partial sums $s_n = u_0 + u_1 + \cdots + u_n$ is summable $E(r)$ to σ if $\sigma_n \to \sigma$ as $n \to \infty$, where

$$E(r): \qquad \sigma_n = \sigma_n(r) = \sum_{k=0}^{n} \binom{n}{k} r^k (1-r)^{n-k} s_k,$$

r being real or complex. In case Σu_n converges, we denote its sum by s. It is well known that $\sigma_n \to s$ for every convergent series, if and only if r lies in the interval $0 < r < 1$; the theorem of this note shows that $\sigma_n \to s$ for specified classes of convergent series, if and only if r lies in correspondingly larger regions of the complex plane.

Theorem.[1] *If $R > 1$, a necessary and sufficient condition for $E(r)$ to have the property that Σu_n is summable $E(r)$ to s whenever $\Sigma u_n z^n$ has its radius of convergence greater than or equal to R, is that*

$$(1) \qquad |r/R| + |1-r| < 1.$$

Proof of sufficiency. Assume that (1) holds, and let $\Sigma u_n z^n$ have radius of convergence $R > 1$, so that Σu_n converges and s is defined. Choose R_1 such that $1 < R_1 < R$, and

$$(2) \qquad |r/R_1| + |1-r| < 1.$$

Then

$$|s - s_n| = \left|\sum_{k=n+1}^{\infty} u_k\right| = \left|\sum_{k=n+1}^{\infty} (u_k R_1^k) R_1^{-k}\right| \leqq M_1 \sum_{k=n+1}^{\infty} R_1^{-k} = M_2 R_1^{-n},$$

* Received September 17, 1945; Revised November 26, 1945.

[1] The author wishes to thank Professor R. P. Agnew for the very simple proof here given.

where $M_1 = \max_{k \geqq 0} | u_k R_1^k |$ and $M_2 = M_1/(R_1 - 1)$. Hence

$$\begin{aligned} | s - \sigma_n | &= \left| \sum_{k=0}^{n} \binom{n}{k} r^k (1-r)^{n-k} (s - s_k) \right| \\ &\leqq \sum_{k=0}^{n} \binom{n}{k} | r |^k | 1 - r |^{n-k} M_2 R_1^{-k} \\ &= M_2 (| r/R_1 | + | 1 - r |)^n, \end{aligned}$$

and therefore (2) implies that $\sigma_n \to s$.

Proof of necessity. Take $u_n = (\theta R)^{-n}$, where $R > 1$ and θ is a complex number, for which $| \theta | = 1$, to be determined later. Then $\Sigma u_n z^n$ has radius of convergence R, and $s = \Sigma u_n = \theta R/(\theta R - 1)$. Also,

$$\begin{aligned} \sigma_n &= \frac{1}{\theta R - 1} \sum_{k=0}^{n} \binom{n}{k} r^k (1-r)^{n-k} [\theta R - (\theta R)^{-k}] \\ &= \frac{\theta R}{\theta R - 1} - \frac{[(r/\theta R) + (1-r)]^n}{\theta R - 1}. \end{aligned}$$

If now we assume that Σu_n is summable $E(r)$ to s, i. e., that $\sigma_n \to s$, it follows that

$$| (r/\theta R) + (1 - r) | < 1,$$

from which (1) is obtained by choosing θ so that $| \theta | = 1$ and $\arg (r/\theta R) = \arg (1 - r)$.

ACADIA UNIVERSITY.

REFERENCE.

1. R. P. Agnew, "Euler transformations," *American Journal of Mathematics*, vol. 66 (1944), pp. 313-338, especially p. 315.

AN ABELIAN LEMMA CONCERNING ASYMPTOTIC EQUILIBRIA.*

By Aurel Wintner.

The original purpose of the following considerations has been a *purely Abelian* approach to the principal limit theorem of [1]. Such an approach requires the elimination of the explicit machinery of the successive approximations, which underly the proof given in [1].

It turns out that such a simplification of the method not only is possible but, corresponding to its more primitive nature, is such as to lead to generalizations which are not within the scope of successive approximations. In fact, the only restriction to be imposed on the function $f = f(t;x)$ defining the (vectorial) differential equation $x' = f(t;x)$ will be a restriction *in the large*, whereas the convergence of the process of successive approximations can be assured only by *local* assumptions (as exemplified by Lipschitz's sufficient condition). In [1], asumptions of *both* kinds were needed. In the sequel, even the *single* assumption to be made, the assumption in the large, is more general than the corresponding assumption of [1].

Another consequence of the omission of any restriction of a local nature will be the inclusion of differential equations in which the solutions are not uniquely determined by the initial conditions. In fact, the question of uniqueness is just as much a purely local affair as the (local) convergence of the process of successive aproximaptions.

Let $f = (f_1, \cdots, f_n)$ and $x = (x_1, \cdots, x_n)$ be vectors with real components f_i, x_i, and let $|u|$ denote the Euclidean length of the vector $u = (u_1, \cdots, u_n)$. Then the generalized Abelian lemma to be proved can be formulated as follows:

Let f be a continuous function of the position $(t;x)$ on the product space of the half-line $0 \leqq t < \infty$ and of the whole Euclidean x-space. Suppose that there exists a pair of functions λ, ϕ satisfying

$$|f(t;x)| < \lambda(t)\phi(|x|), \tag{1}$$

where $\lambda(r)$ and $\phi(r)$ are positive when $0 \leqq r < \infty$, continuous when $0 < r < \infty$, and are subject to

* Received March 22, 1946.

(2) $$\int^{\infty} \lambda(t)\, dt < \infty$$

and

(3) $$\int^{\infty} (dr)/\phi(r) = \infty.$$

Then, if x_0 is any point of the x-space, and if

(4) $$x = x(t), \quad 0 \leqq t < \epsilon = \epsilon(x_0)$$

is any solution vector of

(5) $$x' = f(t;x), \quad x(0) = x_0, \qquad (x' = dx/dt),$$

it is possible to extend (4) *to a solution of* (5) *over the whole half-line* $0 \leqq t < \infty$, *and every solution vector*

(6) $$x = x(t), \quad 0 \leqq t < \infty$$

is such as to tend to a (finite) limit vector, $x(\infty)$, as $t \to \infty$.

This is the more interesting as (for reasons pointed out above) a solution (6) of $x' = f(t;x)$ need not be uniquely determined by the initial condition $x(0) = x_0$. As a matter of fact, the integration constant $x(\infty)$ (which is claimed to exist for every $x(0)$ and *for every choice of* (6) *when $x(0)$ is fixed*) need not be a single-valued function of $x(0)$.

Since (1) remains fulfilled if λ, ϕ are replaced by functions λ^*, ϕ^* satisfying $\lambda \leqq \lambda^*$, $\phi \leqq \phi^*$, and since (2), (3) are restrictions on λ, ϕ only when t, r are large, it is clear that, without violating (2) and (3), it can be assumed that

(7) $$\limsup_{t \to +0} \lambda(t) < \infty$$

and

(8) $$\liminf_{r \to +0} \phi(r) \neq 0, \qquad (\text{and } \phi(0) \neq 0).$$

The positive function $\lambda(t)$ is supposed to be continuous when $0 < t < \infty$ (actually, (2) alone would suffice). In view of (7), this implies that, if $0 < T < \infty$, there exists a constant $\lambda = \lambda_T$ satisfying $\lambda(t) < \lambda_T$ when $0 \leqq t \leqq T$. It follows therefore from (1) and (3) that $\phi_T(r) = \lambda_T \phi(r)$ is a continuous function having the following properties:

$$|f(t;x)| < \phi_T(|x|) \text{ if } 0 \leqq t \leqq T$$

and

$$\int^{\infty} (dr)/\phi_T(r) = \infty.$$

But these properties imply that every solution vector (4) of $x' = f(t; x)$ can be continued over the whole interval $0 \leqq t \leqq T$ (cf. [2], Appendix). Since T can be chosen arbitrarily large, it follows that every solution (4) can be extended into a solution (6). Hence, only the assertion following (6), that is, the existence of $x(\infty)$, remains to be proved.

To this end, let (6) be any solution of $x' = f(t; x)$. Then, from (1),

$$\int_u^v |\, dx(t)|/\phi(|\, x(t)\,|) < \int_u^v \lambda(t)\, dt, \tag{9}$$

if $0 < u < v < \infty$. On the other hand, since $|\, x(t)\,|$ and $\phi(r)$ are real-valued and continuous, and since a real-valued, continuous function cannot leave out a value, it is clear that

$$\int_\alpha^\beta (dr)/\phi(r) \leqq \int_u^v |\, dx(t)\,|/\phi(|\, x(t)\,|), \tag{10}$$

where

$$\alpha = \alpha(u, v) = \min_{u \leqq t \leqq v} |\, x(t)\,|, \quad \beta = \beta(u, v) = \max_{u \leqq t \leqq v} |\, x(t)\,|. \tag{11}$$

Finally, it is seen from (9) and (1) that there exists a t_0 having the property that

$$\int_u^v |\, dx(t)\,|/\phi(|\, x(t)\,|) < 1 \text{ if } v > u > t_0. \tag{12}$$

According to (10) and (12), the functions (11) satisfy the inequality

$$\int_\alpha^\beta (dr)/\phi(r) < 1 \text{ if } v > u > t_0.$$

In view of (3), this inequality implies that both functions (11) remain bounded as $u \rightarrow \infty$ (hence $v \rightarrow \infty$). In order to see this, it is sufficient to let v tend to ∞ while u $(> t_1)$ is fixed, and to observe that both functions (11) are monotone in u and in v.

Accordingly, the second of the functions (11) is bounded on the range

$0 < u < v < \infty$. This means that there exists a constant, say M, satisfying $|x(t)| < M$, where $0 \leqq t < \infty$. On the other hand, (8) and the continuity of the positive function $\phi(r)$ on the half-line $0 < r < \infty$ imply the existence of a positive constant, say m, satisfying $\phi(r) > m$, where $0 \leqq r < M$. Hence it is seen from (12) that

$$\int_u^v |dx(t)| < \text{const. if } v > u > t_0.$$

Accordingly, *the solution vector* (6) *is of bounded variation,*

$$\int^\infty |dx(t)| < \infty.$$

The statement made after (6), namely, the existence of a limiting position for the vector $x(t)$ as $t \to \infty$, is just a corollary of this fact.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES

[1] A. Wintner, "Asymptotic equilibria," *American Journal of Mathematics*, vol. 68 (1946), pp. 125-132.

[2] ———, "The infinities in the non-local existence theorem of ordinary differential equations," *ibid.*, pp. 173-178.

COMPLEX SOLUTIONS OF PARTIAL DIFFERENTIAL EQUATIONS.*

By Herman Chernoff.

1. Introduction. In order to study solutions of partial differential equations

$$L(u) \equiv \Delta u + A(x, y)u_x + B(x, y)u_y + C(x, y)u = 0 \tag{1.1}$$

Bergman considered "classes" of complex solutions of (1.1). Some of these "classes" have useful properties; their introduction frequently helps in the investigation of real solutions in a manner which is analogous to the way analytic functions of a complex variable help in the study of real solutions of the Laplace equation $\Delta u = 0$ (see Bergman, [1], [2], [3]).[1]

Let $E(z, \bar{z}, t)$ [2] denote a complex function of the three real variables x, y and t which is defined in a domain of the (x, y)-plane containing the origin and for $-1 \leq t \leq 1$.

The totality of functions

$$u(z, \bar{z}) = P[f] \equiv \int_{-1}^{1} E(z, \bar{z}, t) f[(z/2)(1-t^2)] \frac{dt}{(1-t^2)^{1/2}} \tag{1.2}$$

which we obtain when $f(\zeta)$ ranges over the totality of analytic functions of a complex variable which are regular at the origin is denoted as the "class" $\mathcal{B}(E)$; E is called the generating function of the class and f is the associate of $u(z, \bar{z})$ with respect to the operator P.

If $E(z, \bar{z}, t)$ is a particular solution of a certain differential equation associated with (1.1), then every function $u(z, \bar{z})$ in $\mathcal{B}(E)$ satisfies (1.1) and thus $\mathcal{B}(E)$ is a "class" of complex solutions of (1.1) (See [1], [2], [3]). Bergman has shown that to every equation (1.1), where A, B, and C are entire functions of x and y, there exists at least one $E(z, \bar{z}, t)$ such that the class $\mathcal{B}(E)$ has the following property: If $U(x, y)$ is a real solution of (1.1) at the origin then there exists a function $u(z, \bar{z})$ in $\mathcal{B}(E)$ such that

* Received November 14, 1945. This paper was prepared while the author was a fellow under the Program of Advanced Research and Instruction in Applied Mathematics at Brown University.

[1] Numbers in brackets refer to bibliography.

[2] The notation $g(z, \bar{z})$ will denote $g_1(x, y) + ig_2(x, y)$ where $z = x + iy$, $\bar{z} = x - iy$ and g_1, g_2, x and y are all real.

$$U(x, y) = Re[u(z, \bar{z})].$$

This is the basic relation between the solutions of the Laplace equation $\Delta u = 0$ and the class $\mathcal{B}(1)$ which is the class of analytic functions regular at the origin.

Various questions arise in connection with the introduction of classes of functions. One of them is the problem of the distribution of the b-points of a function $u(z, \bar{z})$ of the class $\mathcal{B}(E)$. Just as in the classical theory of analytic functions of a complex variable this problem consists mainly of relations between the behavior of

$$m[r, u(z, \bar{z})] = (1/2\pi) \int_0^{2\pi} \log^+ | u(re^{i\phi}, re^{-i\phi}) | \, d\phi \; ^3$$

and

$$n[r, (u(z, \bar{z}) - b)^{-1}] = (1/2\pi i) \int_{|z|=r} d \log [u(z, \bar{z}) - b]$$

for $r \rightarrow \infty$. In addition the relations between these quantities and the coefficients of the series expansion of $u(z, \bar{z})$ at the origin is of importance.

If $u(z, \bar{z}) \neq b$ on $| z | = r$, $n[r, (u(z, \bar{z}) - b)^{-1}]$ represents the "number" of b-points of the function $u(z, \bar{z})$ in $| z | \leq r$. (A more detailed explanation of what is to be understood by "number" will be given in the next section.)

We note further that Bergman showed (See [3, § 8]) that when the E function is of a certain form, it is possible to obtain upper bounds for the growth of $n[r, (u(z, \bar{z}) - b)^{-1}]$ in terms of the coefficients $\{A_{m0}\}$ of the expansion

$$u(z, \bar{z}) = \sum_{m,n=0}^{\infty} A_{mn} z^m \bar{z}^n.$$

It is the object of this paper to derive an upper bound for $m[r, u(z, \bar{z})]$ in terms of $\sum_{\nu=1}^{q} n[r, (u - a_\nu)^{-1}]$, and other quantities. This inequality is analogous to the inequality of the Second Fundamental Theorem of Nevanlinna for meromorphic functions.

These considerations suggest the extension of the Picard Theorem[4] to complex solutions of elliptic equations. It will be seen that under certain conditions an analogue can be established.

In Section **6**, certain results concerning coefficients of the expansion of

---

[3] $\log^+ | a | = \log | a |$ if $| a | \geq 1$
$\log^+ | a | = 0$ if $| a | \leq 1$

[4] In one form the Picard Theorem states that an entire function assumes all finite values with only one possible exception.

"pseudo-simple" functions are derived. These results represent a generalization of classical theorems concerning the coefficients of the expansion of simple (univalent) functions.

2. Generating Functions of a Special Form. In this section we shall describe the general idea of the treatment of certain classes of "entire functions" developed by Bergman in [3] and we shall formulate some previous results which will be used in what follows. Furthermore we shall discuss in detail the properties of certain notions which are of importance in connection with the study of the distribution of b-points.

In Bergman's approach the class of functions

$$u(z,\bar{z}) = \int_{-1}^{1} E(z,\bar{z},t) f[(z/2)(1-t^2)] \frac{dt}{(1-t^2)^{\frac{1}{2}}} \tag{2.1}$$

is of special interest when there exists a function $E^{(1)}(z,s,t)$ which is an entire function of the two complex variables z and s, with the property that if $|z| = s$, then

$$E(z,\bar{z},t) = E^{(1)}(z,s,t). \tag{2.2}$$

If, in addition, we impose the condition that $f(z)$ be entire. we have a class of entire solutions of our elliptic equation such that for $|z| = s$

$$u(z,\bar{z}) = v_s(z) \equiv \int_{-1}^{1} E^{(1)}(z,s,t) f[(z/2(1-t^2)] \frac{dt}{(1-t^2)^{\frac{1}{2}}} \tag{2.3}$$

and $v_s(z)$ is an entire function of z for each value of s. This means that on the circle $|z| = s, u(z,\bar{z})$ takes on the values of an entire function of z. This property will be the one through which we shall be able to apply classical theorems of analytic functions of a complex variable to theorems concerning these classes of complex solutions of elliptic equations.

An example of such a class is the class of solutions of $\Delta u + u = 0$ where $E(z,\bar{z},t) = e^{it(z\bar{z})^{\frac{1}{2}}}$ and hence $E^{(1)}(z,s,t) = e^{ist}$. Thus for $|z| = s$, and $f(z) = \sum_{0}^{\infty} a_n z^n$

$$u(z,\bar{z}) = v_s(z) = \int_{-1}^{1} e^{ist} \sum_{0}^{\infty} a_n [(z/2)(1-t^2)]^n \frac{dt}{(1-t^2)^{\frac{1}{2}}} = \sum_{0}^{\infty} a_n c_n \frac{J_n(s)}{s^n} z^n$$

where $J_n(s)$ is the Bessel Function of order n and c_n are positive constants. We shall return to this example later.

In general if $v_s(z) = \Sigma c_{mn} z^m s^n$, $u(z,\bar{z}) = \Sigma A_{mn} z^m \bar{z}^n$, $(m > n)$, $f(z) = \Sigma a_n z^n$ and $E^{(1)}(z,s,t) = \Sigma E_\mu^{(m)}(t) s^m z^\mu$, then the following relations have been established by Bergman (See [3, § 7])

$$c_{mn} = \sum_{\nu=0}^{n} \frac{A_{\nu 0}\Gamma(\nu+1)}{\pi^{\frac{1}{2}}\Gamma(\nu+\frac{1}{2})} \int_{-1}^{1} E_{n-\nu}^{(m)}(t)\,[1-t^2]^{\nu-\frac{1}{2}}dt, \tag{2.4}$$

$$A_{n0} = \frac{a_n\pi^{\frac{1}{2}}\Gamma(n+\frac{1}{2})}{2^n\Gamma(n+1)}. \tag{2.5}$$

In order to study the distribution of b-points of a function $u(z,\bar{z})$ the following analogue is considered. For an analytic function of a complex variable, $f(z)$, the number of b-points in a region bounded by a set of Jordan curves C on which there are no b-points is

$$d = (1/2\pi i) \int_C d\log[f(z)-b].$$

This expression also accounts for the multiplicity of the b-points. Analogously, we may consider

$$d = (1/2\pi i) \int_C d\log[u(z,\bar{z})-b] \tag{2.6}$$

to characterize the region with respect to the value b of $u(z,\bar{z})$ if $u(z,\bar{z})$ is continuously differentiable along the boundary and continuous in the region (See [5]).

The following properties can be proved, if $u(z,\bar{z})$ is continuously differentiable in x and y.

(i) d is always an integer, either negative, positive or zero. For functions analytic in the domain, however, d is always greater than or equal to zero.

(ii) If d_1 characterizes the region R_1 with respect to the b values of $u(z,\bar{z})$, d_2 characterizes the region R_2 with respect to the b values of $u(z,\bar{z})$, and R_1 and R_2 have no inner points in common, then d_1+d_2 characterizes the region $R_3 = R_1 \cup R_2$ with respect to the b values of $u(z,\bar{z})$.

(iii) If d is not zero then $u(z,\bar{z})$ assumes the value b at least once in the region.

(iv) d is a characteristic of the b-points in the region and does not depend upon the region itself. That is to say, if we deform the region so that no b-points are omitted or added, d remains fixed. Thus every *set of b-points* which may be isolated from other b-points by a region, whose boundary consists of a set of Jordan curves, has an *index*. In particular, an isolated b-point has an index and for an analytic function of a complex variable, the index of a b-point is the multiplicity of the value b at the point.

We may now define

$$n[s,(u(z,\bar{z})-b)^{-1}] = (1/2\pi i) \int_{|z|=s} d\log[u(z,\bar{z})-b]. \tag{2.6}$$

In the classical theory of functions of a complex variable, if $f(z)$ is analytic in $|z| \leq s$, $n[s,(f(z)-b)^{-1}]$ is defined as the number (counting multiplicity) of b-points of $f(z)$ in $|z| \leq s$ if $f(z) \not\equiv b$. Our definition coincides with this if $f(z)$ has no b-points on $|z| = s$. If $f(z)$ has b-points on $|z| = s$ our definition does not apply.[5] In the following we shall write $\tilde{n}[s,(f(z)-b)^{-1}]$ to stand for the number of b-points (counting multiplicity) of the analytic function $f(z)$ in $|z| \leq s$.

For the class of functions in which we are interested $u(z,\bar{z}) = v_s(z)$ for $|z| = s$. Thus, if $n[s,(u(z,\bar{z})-b)^{-1}]$ is defined,

$$(2.7) \qquad n[s,(u(z,\bar{z})-b)^{-1}] = n[s,(v_s(z)-b)^{-1}] = \tilde{n}[s,(v_s(z)-b)^{-1}].$$

Making use of the fact that $\tilde{n} \geq 0$, Bergman proved (See [3, § 8]) that, if it exists,

$$(2.8) \qquad n[s,(u(z,\bar{z})-b)^{-1}] \geq 0.$$

Referring to the example $\Delta u + u = 0$, we had, if $f(z) = z^m$,

$$v_s(z) = c_m \frac{J_m(s)}{s^m} z^m.$$

If $b = 0$, it is evident that if $J_m(s) \neq 0$, $n[s,(u(z,\bar{z})-b)^{-1}]$ is defined and thus

$$n[s, u(z,\bar{z})^{-1}] = \tilde{n}[s,(c_m \frac{J_m(s)}{s^m} z^m)^{-1}] = m.$$

This means that the index of the zero-point at the origin is m. Now, $u(z,\bar{z})$ is also zero on the set of circles of radius s_ν where $J_m(s_\nu) = 0$. The index of each of these circles is zero for, if we consider for the region isolating the circle $|z| = s_\nu$ from all other zero points the ring bounded by $|z| = s_\nu - \epsilon$ and $|z| = s_\nu + \epsilon$, ϵ small, then the index of the circle is $n[s+\epsilon, u(z,\bar{z})^{-1}] - n[s-\epsilon, u(z,\bar{z})^{-1}] = 0$.[6] If $b \neq 0$, we have

$$\begin{aligned} n[s,(u(z,\bar{z})-b)^{-1}] &= m \qquad &&\text{if } |b| < |c_m J_m(s)| \\ n[s,(u(z,\bar{z})-b)^{-1}] &= 0 &&\text{if } |b| > |c_m J_m(s)|. \end{aligned}$$

On each circle for which $|b| = |c_m J_m(s)|$, there are m b-points and it can be shown without trouble that the index of each b-point is $+1$ or -1. Also, on the first such circle the indices are each $+1$, on the second -1, on the

[5] It is noteworthy that if $f(z)$ is analytic and has b-points on $|z| = s$, then although our definition does not apply, the integral of our definition will give the number of b-points in $|z| < s$ plus half the number on $|z| = s$ (counting multiplicity).

[6] The convention of integrating in the direction such that the region is to the left must be observed and gives the negative sign to $n[s-\epsilon, u(z,z)^{-1}]$.

third $+1$, alternating in this manner until $J_m(s)$ becomes so small that there are no more of these circles.

Geometrically speaking, if an isolated b-point has an index of -1, it means that as a small curve is drawn counter clockwise about the b-point in the z plane, the corresponding $u(z,\bar{z})$ points describe a curve drawn clockwise about b in the u plane. An index of -2 would mean that the $u(z,\bar{z})$ points describe a path which circles $u=b$ twice in a clockwise direction.

In addition, a concept used in the theory of meromorphic functions is very easily generalized. If $f(z)$ is a meromorphic function then there is defined

$$m[s,f(z)] = (1/2\pi) \int_0^{2\pi} \log^+ |f(se^{i\theta})|\, d\theta.$$

Analogously, given a function $g(z,\bar{z})$, one defines (See Bergman [3, § 8])

$$m[s,g(z,\bar{z})] = (1/2\pi) \int_0^{2\pi} \log^+ |g(se^{i\theta}, se^{-i\theta})|\, d\theta \tag{2.9}$$

if the integral exists. If $g(z,\bar{z}) = f(z)$ on $|z|=s$ where $f(z)$ is meromorphic, then $m[s,g(z,\bar{z})]$ exists. Thus $m[s,u(z,\bar{z})]$ exists and if $u(z,\bar{z}) \not\equiv b$ on $|z|=s$, then $m[s,(u(z,\bar{z})-b)^{-1}]$ exists too. In addition

$$m[s,u(z,\bar{z})] = m[s,v_s(z)]; \quad m[s,(u(z,\bar{z})-b)^{-1}] = m[s,(v_s(z)-b)^{-1}]. \tag{2.10}$$

We may consider $m[s,(u(z,z)-b)^{-1}]$ to be a measure of the intensity of the values b near the circle $|z|=s$. Furthermore, since $u(z,\bar{z})$ is always finite, $m[s,u(z,\bar{z})]$ may be considered as a measure of the growth of the function $u(z,\bar{z})$.

Using the first Theorem of Nevanlinna, Bergman derived the following bounds (see [3, § 8]):
If $u(z,\bar{z}) = \Sigma A_{mn} z^m \bar{z}^n$ and

$$\rho = \limsup_{m+n\to\infty} \frac{(m+n)\log(m+n)}{\log \left| 1/\sum_{\nu=0}^{n} \dfrac{A_{\nu 0}\Gamma(\nu+1)}{\pi^{1/2}\Gamma(\nu+\frac{1}{2})} \displaystyle\int_{-1}^{1} E_{n-\nu}^{(m)}(t)(1-t^2)^{\nu-\frac{1}{2}} dt \right|} < \infty$$

then for arbitrary $\epsilon > 0$

(a) $\max u(z,\bar{z}) < c_1 e^{s^{\rho+\epsilon}}$, $\therefore m[s,u(z,\bar{z})] \leq s^{\rho+\epsilon}$ for s large enough

(b) $n[s,(u(z,\bar{z})-b)^{-1}] < c_2 s^{\rho+\epsilon}$ if $|v_s(0)-b| \geq c > 0$ where c is an arbitrary fixed constant.

(c) $m[s,(u(z,\bar{z})-b)^{-1}] < c_3 s^{\rho+\epsilon}$ if $|v_s(0)-b| \geq c$.

3. An Upper Bound for $m[s, u(z, \bar{z})]$ in terms of $n[s, (u(z, \bar{z}) - \alpha_\nu)^{-1}]$. In this section, we shall apply a modification of the second Fundamental Theorem of Nevanlinna to obtain an upper bound for $m[s, u(z, \bar{z})]$ which will depend largely on $n[s, (u(z, \bar{z}) - \alpha_\nu)^{-1}]$ for q values of α_ν. The modification of the Second Fundamental Theorem will be proved in the Appendix, because the argument involved is not inherently useful in our considerations. The theorem may be stated as follows:

THEOREM 1. *If $f(z) = c_0 + c_k z^k + \cdots$ is an entire function of z, c_0 and c_k not zero, and if $\alpha_1, \alpha_2, \cdots, \alpha_q$ are distinct finite complex numbers and $\tilde{n}[r_1, (f(z) - \alpha_\nu)^{-1}] = 0$, then for $r > r_1$ and $\rho > r$*

$$(3.1) \quad (q-1) m[r, f(z)] \leq [\log (r/r_1)[\sum_{\nu=1}^{q} \tilde{n}[r, (f - \alpha_\nu)^{-1}]] - N_1(r) + S_1(r)$$

where (i) $N_1(r) \geq 0$ *for* $r \geq 1$

$$(3.2) \quad \text{(ii)} \quad \begin{aligned} S_1(r) = 56 &+ [\log^+ |(1/kc_k)| + 6q \log^+ R_c] \\ &+ (q+4) \log^+ R + 2q \log (2q/\delta) \\ &+ q \log 2 + 4 \log^+ (1/r) + 8 \log^+ \rho + 6 \log^+ (1/\rho - r) \\ &+ [8 \log^+ m(\rho, f) + \sum_{\nu=1}^{q} \log | c_0(r) - \alpha_\nu |] \end{aligned}$$

(iii) $\delta = \min [| \alpha_n - \alpha_k |, 1] \qquad (n \neq k)$

$R = \max | \alpha_\nu |$

$R_0 = \max \{|(1/c_0)|, |1/(c_0 - \alpha_\nu)|\}$

and

THEOREM 1. (a): *If $f(z) = \sum_{n=\lambda}^{\infty} c_n z^n$ is an entire function of z, $\gamma > 0$, and $c_\lambda \neq 0$ and if $\alpha_1, \alpha_2, \cdots, \alpha_q$ are distinct finite complex numbers, and $\tilde{n}[r_1, (f(z) - \alpha_\nu)^{-1}] = 0$; then for $r > r_1$ and $\rho > r$, (3.1) holds, where*

(i) $N_1(r) \geq 0$ *for* $r \geq 1$

$$(3.2a) \quad \text{(ii)} \quad \begin{aligned} S_1(r) = 66 &+ [\log^+ |(1/\lambda c_\lambda)| + 6q \log^+ R_0] \\ &+ (4 + 2q) \log^+ R + 2q \log (2q/\delta) + q \log 2 \\ &+ 9 \log^+ (1/r) + 8 \log^+ \rho + 6 \log^+ (1/\rho - r) \\ &+ [8 \log^+ m(\rho, f)] + 5 \log^+ | \lambda | \end{aligned}$$

(iii) *δ, R are defined as in Theorem 1:*

$R_0 = \max \{|(1/c_\lambda)|, |(1/\alpha_\nu)|\}$.

It is of great importance that all terms in $S_1(r)$ which are not in the brackets are independent of the particular function $f(z)$. It is also note-

worthy that if $\alpha_\nu = 0$ in Theorem 1a or $\alpha_\nu = c_0$ in Theorem 1, then R_0 becomes infinite and the inequality (3.2) is useless. This difficulty may be overcome but to do so will be of no value to us because as $c_0 \to \alpha_\nu$, $R_0 \to \infty$ and in our considerations c_0 will not be fixed.

Now, $v_s(z) = \sum_{m=\lambda}^{\infty} c_m(s) z^m$ where $c_m(s)$ is an entire function of s and $c_\lambda(s)$ is the first coefficient which does not vanish identically. If $\lambda = 0$, then suppose that there is a k such that $c_k(s)$ is the next coefficient which does not vanish identically. This is equivalent to assuming that $v_s(z) = c_0(s) + c_k(s) z^k + \cdots$ is not a constant for each fixed s. Because an entire function which is not identically zero can vanish only on an enumerable set of points, one of the following two alternatives holds; except for an enumerable set of values of s along the positive real axis either:

Case 1. $v_s(z) = c_0(s) + c_k(s) z^k + \cdots$ where $c_0(s) \neq 0$ and $c_k(s) \neq 0$

or Case 1a. $v_s(z) = c_\lambda(s) z^\lambda + \cdots$ where $c_\lambda(s) \neq 0$ and $\lambda > 0$.

Thus we may apply Theorem 1 to $f(z) = v_s(z)$ where s is not an element of the above mentioned enumerable set. We obtain

$$(3.3) \quad (q-1) m[r, v_s(z)] \leq [\log(r/r_1)] \sum_{\nu=1}^{q} \bar{n}[r_1, (v_s(z) - \alpha_\nu)^{-1}] + S(r, s)$$

where $r_1 = r_1(s)$, $r > r_1(s)$ and $S(r,s)$ is $S_1(r) - N_1(r)$ corresponding to $f(z) = v_s(z)$.

It was pointed out that as $c_0(s) \to \alpha_\nu$, $R_0 \to \infty$, R_0 is a function of s and it is important for the sake of our inequality to have R_0 bounded. It is likewise important for us to have $|1/c_\lambda|$ or $|1/c_k|$ bounded depending on the case under consideration. On the other hand R and δ are independent of s. Our method of attack will be to find that set of values of s for which R_0 and $|(1/c_\lambda)|$ or $|(1/c_k)|$ are bounded. For this purpose, let us define the following sets of values of s along the positive real axis.

$\mathfrak{S}_1(b, c)$	= the set of points for which	$\lvert v_s(0) - b \rvert \geq c$ i. e. $\lvert c_0(s) - b \rvert \geq c$	
$\mathfrak{S}_2(b, c)$	= " " " " " "	$\lvert v_s(0) - b \rvert \leq c$ " $\lvert c_0(s) - b \rvert \leq c$	
$\mathfrak{S}_3$	= " " " " " "	$c_0(s) \neq 0$ and $c_k(s) \neq 0$	[in case 1]
$\mathfrak{S}_3$	= " " " " " "	$c_\lambda(s) \neq 0$	[in case 1a]
$\mathfrak{S}_4(b, r_1)$	= " " " " " "	$v_s(z) \neq b$ for $\lvert z \rvert \leq r_1$	
$\mathfrak{S}_5(0, c)$	= " " " " " "	$\lvert c_k(s) \rvert \geq c$	[in case 1]
$\mathfrak{S}_6(0, c)$	= " " " " " "	$\lvert c_\lambda(s) \rvert \geq c$	[in case 1a]

From these definitions it follows that if $c_0(s) \neq 0$, and $s \in \mathfrak{S}_1(\alpha_\nu, c)$

$\cap \mathfrak{S}_1(0, c)$ for $\nu = 1, 2, \cdots q$, then $R_0 \leq 1/c$. Furthermore if $s \varepsilon \mathfrak{S}_5(0, c)$, $|1/c_k| \leq 1/c$. In addition, $N_1(r) \geq 0$ for $r \geq 1$. Thus if $c_0(s) \neq 0$ and $c_k(s) \neq 0$, i.e. $s \varepsilon \mathfrak{S}_3$, we may apply Theorem 1 and the definition of $S(r, s)$ to state:

THEOREM 2: *If*

(i) $c_0(s) \not\equiv 0$, $c_k(s) \not\equiv 0$

(ii) $s \varepsilon \mathfrak{S}_1(0, c) \cap \mathfrak{S}_1(\alpha_\nu c) \cap \mathfrak{S}_5(0, c) \cap \mathfrak{S}_3$ $\qquad \nu = 1, 2, \cdots q$

(iii) $r \geq 1$

(iv) $\rho > r$

Then there is a constant $c_1[k, c, q, \alpha_\nu]$ *such that*

$$S(r, s) \leq c_1[k, c, q, \alpha_\nu] + 4 \log^+ (1/r) + 6 \log^+ (1/\rho - r) + 8 \log^+ \rho + 8 \log^+ m[\rho, v_s(z)] + \sum_{\nu=1}^{q} \log | c_0(s) - \alpha_\nu | . \tag{3.4}$$

In addition, we have by a similar argument

THEOREM 2a: *If*

(i) $c_\lambda(s) \not\equiv 0$. $\lambda > 0$

(ii) $s \varepsilon \mathfrak{S}_1(\alpha_\nu, c) \cap \mathfrak{S}_6(0, c) \cap \bar{\mathfrak{S}}_3$

(iii) $r \geq 1$

(iv) $\rho > r$

Then there is a constant $c_2[\lambda, c, q, \alpha_\nu]$ *such that*

$$S(r, s) \leq c_2[\lambda, c, q, \alpha_\nu] + 9 \log^+(1/r) + 6 \log^+(1/\rho - r) + 8 \log^+ \rho + 8 \log^+ m[\rho, v_s(z)] . \tag{3.4a}$$

Now, we have from (2.10)

$$m[s, u(z, \bar{z})] = m[s, v_s(z)]$$

and furthermore if there are no α-points of $u(z, \bar{z})$ on $|z| = s$

$$n[s,(u(z, \bar{z}) - \alpha)^{-1}] = n[s,(v_s(z) - \alpha)^{-1}] = \tilde{n}[s,(v_s(z) - \alpha)^{-1}] .$$

In addition, we have if $s \varepsilon \mathfrak{S}_4(\alpha_\nu, \tilde{r})$, $v_s(z) \neq \alpha_\nu$ for $|z| \leq \tilde{r}$. Hence $\tilde{n}[\tilde{r},(v_s(z) - \alpha_\nu)^{-1}] = 0$. Thus if $s \varepsilon \mathfrak{S}_4(\alpha_\nu, \tilde{r})$ $\nu = 1, 2, \cdots q$, then $r_1(s) \geq \tilde{r}$. We may now apply (3.3) and Theorems 2 and 2a to the case where $r = s$, whence we obtain:

THEOREM 3: *If*

(i) $c_0(s) \not\equiv 0$, $c_k(s) \not\equiv 0$

(ii) $s \in \mathfrak{S}_1(0, c) \cap \mathfrak{S}_1(\alpha_\nu, c) \cap \mathfrak{S}_5(0, c) \cap \mathfrak{S}_3 \cap \mathfrak{S}_4(\alpha_\nu, \tilde{r})$, $\quad \nu = 1, 2, \cdots q$

(iii) $s \geq 1$, *and* $s > \tilde{r}$

(iv) *There are no* α_ν*-points of* $u(z, \bar{z})$ *on* $|z| = s$

(v) $\rho > s$

Then

$$(3.5) \quad (q-1)m[s, u(z, \bar{z})] \leq [\log s/\tilde{r}][\sum_{\nu=1}^{q} n[s, (u-\alpha_\nu)^{-1}] + 8 \log^+ \rho + 6 \log^+ (1/\rho - s) + 8 \log^+ m[\rho, v_s(z)] + c_1[k, c, q, \alpha_\nu] + \sum_{\nu=1}^{q} \log | c_0(s) - \alpha_\nu |$$

and we also have

THEOREM 3a: *If*

(i) $c_\lambda(s) \not\equiv 0$, $\quad \lambda > 0$

(ii) $s \in \mathfrak{S}_1(\alpha_\nu, c) \cap \mathfrak{S}_6(0, c) \cap \bar{\mathfrak{S}}_3 \cap \mathfrak{S}_4(\alpha_\nu, \tilde{r})$ $\quad \nu = 1, 2, \cdots q$

(iii), (iv), (v) *same as in Theorem 3*

Then

$$(3.5a) \quad (q-1)m[s, u(z, \bar{z})] \leq [\log (s/\tilde{r})][\sum_{\nu=1}^{q} n[s, (u-\alpha_\nu)^{-1}]] + 8 \log^+ \rho + 6 \log^+ (1/\rho - s) + 8 \log^+ m[\rho, v_s(z)] + c_2[\lambda, c, q, \alpha_\nu].$$

The inequalities (3.5) and (3.5a) are the inequalities we desired. It is to be noted that the only restriction on ρ is that $\rho > s$. A special case of these inequalities which is of great importance is the case where $\rho = \mu s$, μ constant and greater than 1. Then the term $\log^+ 1/\rho - s$ is bounded and may be incorporated into c_1 or c_2. In addition $\log^+ \rho \leq \log^+ \mu + \log^+ s$ and $\log^+ \mu$ may also be incorporated into c_1 or c_2.

It is also noteworthy that the above inequalities are quite general in the sense that they may apply to *any* function $u(z, \bar{z})$ which takes on the values of $v_s(z)$ for $|z| = s$ where $v_s(z)$ is entire in z and s. These inequalities can now be adapted to the study of classes of solutions of *particular*

partial differential equations by translating the hypotheses of Theorems 3 and 3a into hypotheses concerning the nature of $E^{(1)}(z, s, t)$.

Before we prove such a theorem involving $E^{(1)}(z, s, t)$, let us consider the following example.

Example.

$\Delta u - u = 0$; Here $E(z, \bar{z}, t) = e^{t(z\bar{z})^{\frac{1}{2}}}$ (see [2])

Hence $E^{(1)}(z, s, t) = e^{st}$ and,

$$v_s(z) = \int_{-1}^{1} e^{st} \frac{f[(z/2)(1-t^2)]}{(1-t^2)^{\frac{1}{2}}} dt.$$

Suppose that

$f(z) = a_0 + a_k z^k + \cdots$ is entire, $a_0 \neq 0$, $a_k \neq 0$.

Then there exists an $\tilde{r}$ such that for $|z| \leq \tilde{r}$

$$f(z) = a_0[1 + \epsilon_1 e^{i\delta_1}] \quad \text{where} \quad |\epsilon_1| < \tfrac{1}{2}.$$

Thus for $|z| \leq \tilde{r}$

$$v_s(z) = a_0 \rho_0(s)[1 + \epsilon_2 e^{i\delta_2}] \tag{3.6}$$

where $|\epsilon_2| < \frac{1}{2}$ and $\rho_0(s) = \int_{-1}^{1} e^{st}(1-t^2)^{-\frac{1}{2}} dt$.

Furthermore

$v_s(z) = c_0(s) + c_k(s) z^k + \cdots$ where

$$c_0(s) = a_0 \int_{-1}^{1} e^{st}(1-t^2)^{-\frac{1}{2}} dt = a_0 \rho_0(s) \tag{3.7}$$

$$c_k(s) = a_k \int_{-1}^{1} (1/2^k) e^{st}(1-t^2)^{k-\frac{1}{2}} dt = a_k \rho_k(s). \tag{3.8}$$

Since $\rho_0(s)$ and $\rho_k(s)$ tend to infinity as $s \to \infty$, (3.6), (3.7) and (3.8) imply that for s large enough

$$s \in \mathfrak{S}_1(0, c) \cap \mathfrak{S}_1(\alpha_\nu, c) \cap \mathfrak{S}_5(0, c) \cap \mathfrak{S}_3 \cap \mathfrak{S}_4(\alpha_\nu, \tilde{r}), \qquad \nu = 1, 2, \cdots, q.$$

Now suppose that

$$\operatorname*{Max}_{|z|=\mu s} |v_s(z)| = O[e^{s^{\nu+\epsilon}}] \qquad \text{for any positive } \epsilon.$$

Then $\log^+ m[\mu s, v_s(z)] \leq \log^+ \log^+ [e^{s^{\nu+\epsilon}}] \leq (\nu + \epsilon) \log^+ s$.

Furthermore $\log |c_0(s) - \alpha_\nu| \leq c + s$.

Thus, if we apply Theorem 3, (3.5) gives

$$(3.9)\qquad (q-1)m[s,u(z,\bar{z})] \leq [\log s/\tilde{r}][\sum_{\nu=1}^{q} n[s,(u-\alpha_\nu)^{-1}]] + qs + 8(1+\nu+\epsilon)\log s$$

for s large enough and such that $u(z,\bar{z}) \neq \alpha_\nu$ on $|z| = s$. Hence if

$\sum_1^q n[s,(u-\alpha_\nu)^{-1}] = O[s^{\delta'+\epsilon}]$ for any positive ϵ, and $\delta = \max[\delta', 1]$,

$$(3.10)\qquad m[s,u(z,\bar{z})] = O[s^{\delta'+\epsilon}]$$

for any positive ϵ and for the set of positive s for which $u(z,\bar{z}) \neq \alpha_\nu$. If this set of s is everywhere dense as in the case where all α_ν points of $u(z,\bar{z})$ are isolated points, then (3.10) holds for the entire set of positive s by continuity.

(3.10) shows that $\sum_{\nu=1}^{q} n[s,(u-\alpha_\nu)^{-1}]$ determines an upper bound for $m[s,u(z,\bar{z})]$ unless $m[s,u(z,\bar{z})] = O[s^{1+\epsilon}]$.

The result of this example can be generalized to prove the following

THEOREM 4. *If*

(i) *There exists an $r_2, \bar{s}, c$, such that for $|z| \leq r_2$, $s > \bar{s}$*

$$E^{(1)}(z,s,t) = \tilde{E}_0(s,t)[1+\epsilon_1 e^{i\delta_1}] \quad \textit{where} \quad |\epsilon_1| \leq c < 1$$

and $\tilde{E}_0(s,t) = E^{(1)}(0,s,t)$ is real and bounded from below for $s > 0$

(ii) $E^{(1)}(z,s,t)f[(z/2)(1-t^2)] = a_0\tilde{E}_0(s,t) + z^k\alpha_k(s,t) + \cdots$
$a_0 \neq 0,\ \alpha_k(s,t) \not\equiv 0$

(iii) $\rho_0(s) = \int_{-1}^{1} \tilde{E}_0(s,t)(dt/(1-t^2)^{1/2}) \to \infty \quad as \quad s \to \infty$

$$\bar{\rho}_k(s) = \left|\int_{-1}^{1} \alpha_k(s,t)(1-t^2)^{-\frac{1}{2}}dt\right| \to \infty \quad as \quad s \to \infty$$

(iv) $h(r) = \max_{|z|=r} |f(z)|$

$$\rho(s) = \max_{|z|=2s} |E^{(1)}(z,s,t)|$$

Then, there is an $\tilde{r} > 0$ such that

$$(3.11)\qquad (q-1)m[s,u(z,\bar{z})] \leq [\log s/\tilde{r}][\sum_1^q n[s,(u-\alpha_\nu)^{-1}]] + 8\log^+\log^+[\rho(s)h(s)] + q\log\rho_0(s) + 8\log s + c$$

for s large enough and such that

$$u(z,\bar{z}) \neq \alpha_\nu \textit{ on } |z| = s.$$

Proof. $f(z)=\sum_0^\infty a_n z^n$, $a_0 \neq 0$. Thus there is an r_3 such that $f(z)=a_0(1+\epsilon_3 e^{i\delta_3})$ where $|\epsilon_3|<(1-c)/16$ for $|z|\leq r_3$. Now $\tilde{r}=\min(r_2, r_3)$. Then

$$v_s(z)=\int_{-1}^{1}\tilde{E}_0(s,t)a_0[1+\epsilon_4 e^{i\delta_4}](dt/(1-t^2)^{1/2})$$

where $|\epsilon_4|<1-(14/16)(1-c)<1$ for $|z|\leq\tilde{r}$, $s>\tilde{s}$. Now, since $\tilde{E}_0(s,t)$ is bounded from below for $s>0$, there exists an M_1 such that

$$E_0(s,t)=-M_1+H(s,t) \text{ where } H(s,t)>0 \text{ for } s>0.$$

Using the fact that $\rho_0(s)\to\infty$, we have

$$\gamma_0(s)=\int_{-1}^{1}(H(s,t)/(1-t^2)^{1/2})\,dt \qquad \text{as } s\to\infty.$$

Indeed $\gamma_0(s)/\rho_0(s)\to 1 \quad$ as $s\to\infty$.

$$\text{Now } v_s(z)=-a_0M_1\int_{-1}^{1}\frac{(1+\epsilon_4 e^{i\delta_4})}{(1-t^2)^{1/2}}\,dt+\int_{-1}^{1}\frac{H(s,t)a_0(1+\epsilon_4 e^{i\delta_4})}{(1-t^2)^{1/2}}\,dt$$

$$=a_0N_1+a_0\gamma(s)[1+\epsilon_5 e^{i\delta_5}]$$

$$\text{where } |N_1|\leq 2M_1\int_{-1}^{1}(dt/(1+t^2)^{1/2})$$

and $|\epsilon_5|\leq 1-(14(1-c)/16)$ for s large enough and $|z|\leq\tilde{r}$.

Thus since $\gamma_0(s)\to\infty$

$v_s(z)=a_0\gamma_0(s)[1+\epsilon_6 e^{i\delta_6}]$ where $|\epsilon_6|\leq 1-(12/16)(1-c)$ for $|z|\leq\tilde{r}$, s large enough or

(3.12) $\quad v_s(z)=a_0\rho_0(s)[1+\epsilon_7 e^{i\delta_7}]$ where $|\epsilon_7|\leq 1-(10/16)(1-c)$ for $|z|\leq\tilde{r}$, s large enough.

Furthermore

(3.13) $\quad |c_0(s)|=|a_0\rho_0(s)|\to\infty$

(3.14) $\quad |c_k(s)|=|\tilde{\rho}_k(s)|\to\infty.$

Thus for s large enough $s\in\mathfrak{S}_1(0,c)\cap\mathfrak{S}_1(\alpha_\nu,c)\cap\mathfrak{S}_5(0,c)\cap\mathfrak{S}_3\cap\mathfrak{S}_4(\alpha_\nu,\tilde{r})$. And for $|z|=2s$

$$|v_s'(z)|=\left|\int_{-1}^{1}E^{(1)}(z,s,t)f[(z/2)(1-t^2)](dt/(1-t^2)^{1/2}\right|$$

$$\leq\rho(s)h(s)\int_{-1}^{1}dt/(1-t^2)^{1/2}.$$

Hence

$$\log^+ m[2s, v_s(z)] \leq \log^+ \log^+ \rho(s) h(s) + \log\log \pi$$

and

$$\log | c_0(s) - \alpha_\nu | \leq c + \log | c_0(s) | \leq c^1 + \log \rho_0(s) .$$

Applying Theorem 3 where $\rho = 2s$, we have for all s large enough and such that $u(z, \bar{z}) \neq \alpha_\nu$ on $| z | = s$, (3.11) which is what we wished to prove. If we know that $\operatorname{Max}_{|z|=2s} | v_s(z) | = w(s)$, it is evident that in (3.11) $\log^+ \log^+ h(s)\rho(s)$ may be replaced by $\log^+ \log^+ w(s)$.

In Theorem 4 the inequality for $m[s, u(z, \bar{z})]$ is stated with reference to certain properties of $E^{(1)}(z, s, t)$. In **5**, certain classes of partial differential equations with known generating functions will be investigated with reference to Theorem 4.

4. An Analogue of the Picard Theorem. The following statement is a simple corollary of Rouché's Theorem: If $f(z)$ is analytic in the circle $| z | \leq r$, $f(0) = 0$ and $| f(z) | \geq | \alpha |$ on $| z | = r$ then $f(z)$ assumes the value α at least once in the circle $| z | < r$. This corollary provides a method of proving an analogue of the Picard Theorem for certain sets [7] of complex solutions of elliptic equations. The method of attack will first be illustrated by an example.

Example. $\Delta u - u = 0$ $\quad E^{(1)}(z, s, t) = e^{st}$. Let $f(z) = \sum_{n=\lambda}^{\infty} a_n z^n$ where $a_\lambda \neq 0$, $\lambda > 0$. Then, there exists an $\tilde{r}$ such that for $| z | \leq \tilde{r}$, $f(z) = a_\lambda z^\lambda [1 + \epsilon_1 e^{i\delta_1}]$, $| \epsilon_1 | < \frac{1}{2}$, so that for $| z | \leq \tilde{r}$

$$v_s(z) = (1/2^\lambda) \int_{-1}^{1} e^{st} a_\lambda z^\lambda [1 + \epsilon_1 e^{i\delta_1}] (1 - t^2)^{\lambda - \frac{1}{2}} dt = a_\lambda z^\lambda \rho_\lambda(s) [1 + \epsilon_2 e^{i\delta_2}]$$

where $| \epsilon_2 | < \frac{1}{2}$ and where $\rho_\lambda(s) = (1/2^\lambda) \int_{-1}^{1} e^{st} (1 - s^2)^{\lambda - \frac{1}{2}} dt \to \infty$ as $s \to \infty$. Thus for $| z | \doteq \tilde{r}$

$$| v_s(z) | \geq \tfrac{1}{2} | a_\lambda | \tilde{r}^\lambda \rho_\lambda(s) \to \infty . \tag{4.1}$$

Consider any finite complex number α and suppose that $u(z, \bar{z})$ does not attain the value α. Then $n[s, (u - \alpha)^{-1}]$ is defined and

$$n[s, (u(z, \bar{z}) - \alpha)^{-1}] = \tilde{n}[s, (v_s(z) - \alpha)^{-1}] = 0 .$$

Now $v_s(0) = 0$ and for $s > \tilde{r}$ and s large enough and $| z | = \tilde{r}$, $| v_s(z) | > | a |$. Thus $v_s(z)$ asumes the value α somewhere in $| z | \leq \tilde{r}$. Hence

---

[7] These sets are subsets of certain classes of complex solutions.

$$n[s,(u(z,\bar{z})-\alpha)^{-1}]=\tilde{n}[s,(v_s(z)-\alpha)^{-1}]\geq\tilde{n}[\tilde{r},(v_s(z)-\alpha)^{-1}]\geq 1$$

and our original asumption that α is not attained by $u(z,\bar{z})$ is contradicted. This means that $u(z,\bar{z})$ assumes all finite values if the associate function $f(z)$ vanishes at the origin.

The above argument suggests the following theorem.

THEOREM 5. *If*

(i) *There exists an* r_2, $\tilde{s}$, c, *such that for* $|z|\leq r_2$, $s>\tilde{s}$

$$E^{(1)}(z,s,t)=\tilde{E}_0(s,t)[1+\epsilon_1 e^{i\delta_1}] \quad where \quad |\epsilon_1|\leq c<1$$

and $\tilde{E}_0(s,t)=E^{(1)}(0,s,t)$ *is real and bounded from below for* $s>0$

(ii) $f(z)=\sum_{n=\lambda}^{\infty}a_n z^n \qquad a_\lambda\neq 0, \quad \lambda>0$

(iii) $\rho_\lambda(s)=(1/2^\lambda)\displaystyle\int_{-1}^{1}\tilde{E}_0(s,t)(1-t^2)^{\lambda-\frac{1}{2}}dt\to\infty \qquad as\ s\to\infty.$

Then all finite values are assumed by $u(z,\bar{z})$.

Proof. Since $f(z)=\sum_{n=\lambda}^{\infty}a_n z^n$ $a_\lambda\neq 0$, $\lambda>0$, there exists an r_3 such that for $|z|\leq r_3$

$$f(z)=z^\lambda a_\lambda[1+\epsilon_3 e^{i\delta_3}] \qquad \text{where } |\epsilon_3|\leq(1/16)(1-c).$$

Following through the argument used to get (3.12) in the proof of Theorem 4, we obtain

$$\begin{aligned}(4.2)\qquad v_s(z)&=z^\lambda a_\lambda[1+\epsilon_7 e^{i\delta_7}]\left[\int_{-1}^{1}(1/2^\lambda)\tilde{E}_0(s,t)(1-t^2)^{\lambda-\frac{1}{2}}dt\right]\\&=z^\lambda a_\lambda\rho_\lambda(s)[1+\epsilon_7 e^{i\delta_7}]\end{aligned}$$

where $|\epsilon_7|<1-(10/16)(1-c)$ for $|z|\leq\tilde{r}=\min[r_2,r_3]$, s large enough. Thus we have

(i) $v_s(0)=0$

(ii) $|v_s(z)|\geq\tilde{r}^\lambda|a_\lambda|\rho_\lambda(s)(10/16)(1-c)$ for s large enough and $|z|=\tilde{r}$. Now given any value α, if $u(z,\bar{z})$ does not attain the value α, then $n[s,(u-\alpha)^{-1}]$ is defined and

$$n[s,(u(z,\bar{z})-\alpha)^{-1}]=\tilde{n}[s,(v_s(z)-\alpha)^{-1}]=0.$$

But for s large enough, $s>\tilde{r}$ and $|z|=\tilde{r}$, $|v_s(z)|>|\alpha|$. Thus $v_s(z)$ assumes the value α in $|z|\leq\tilde{r}$. Hence

$$n[s,(u(z,\bar{z})-\alpha)^{-1}]=\tilde{n}[s,(v_s(z)-\alpha)^{-1}]\geq\tilde{n}[\tilde{r},(v_s(z)-\alpha)^{-1}]\geq 1$$

and this contradiction proves our theorem.

In Theorems 4 and 5 the conditions on $E^{(1)}(z,s,t)$ may be relaxed somewhat by the following consideration. Suppose that $E^{(2)}(z,s,t)$ satisfies the conditions of these theorems and that

$$E^{(1)}(z,s,t)=g[(z/2)(1-t^2)]E^{(2)}(z,s,t)$$

where $g(\zeta)$ is an entire function of ζ. Then it is evident that the class of entire functions generated by $E^{(1)}$ is included in the class generated by $E^{(2)}$. Thus the theorems hold for the class generated by $E^{(1)}$ because they hold for the class generated by $E^{(2)}$.

It is worth mentioning that Theorem 3 suggests another attack on the analogue of the Picard Theorem. If we suppose that there are two finite complex numbers α_1 and α_2 which are not attained by $u(z,\bar{z})$, we have

$$n[s,(u(z,\bar{z})-\alpha_1)^{-1}]=n[s,(u(z,\bar{z})-\alpha_2)^{-1}]=0.$$

Then Theorem 3 states that for a certain set of positive values of s for which conditions (ii), (iii) and (iv) are satisfied.

$$\text{(4.3)}\qquad m[s,u(z,\bar{z})]<8\log^+\rho+8\log^- m[\rho,v_s(z)]+6\log^+(1/\rho-s)+\sum_{\nu=1}^{2}\log|c_0(s)-\alpha_\nu|+c.$$

If we could prove that there is a set of values of s for which (ii), (iii) and (iv) are satisfied and which has elements tending to ∞, and there exists an $h(s)\to\infty$ such that

$$\text{(4.4)}\qquad m[s,u(z,\bar{z})]\geq 8\log\rho+8\log^+ m[\rho,v_s(z)]+6\log^+(1/\rho-s)+\sum_{\nu=1}^{2}\log|c_0(s)-\alpha_\nu|+h(s)$$

the resulting contradiction would imply that $u(z,\bar{z})$ assumes one of the two values α_1 and α_2.

5. Examples. Thus far, all theorems which refer to particular partial differential equations refer to them in terms of $E^{(1)}(z,s,t)$ only. It would be of great value to be able to state theorems in terms of the coefficients of the partial differential equations.

Bergman (See [2, § 3] and see also [7]) has given a method of determining E functions corresponding to equations with certain coefficients. Suppose that the equation has the form

(5.1) $$(\Delta u/4) + A(x,y)u_x + B(x,y)u_y + C(x,y)u = 0$$
or
(5.1′) $$u_{z\bar z} + a(z,\bar z)u_z + \overline{a(z,\bar z)}u_{\bar z} + c(z,\bar z)u = 0$$
where we define
$$f_z = \tfrac{1}{2}[f_x - if_y] \qquad f_{\bar z} = \tfrac{1}{2}[f_x + if_y]\ [8].$$
Then if (i) $a_{\bar z} = \bar a_z$

(ii) $F_z = 0$ where $F = -a_z - |a|^2 + c(z,\bar z)$

it follows that

$E(z,\bar z,t) = \exp P(z,\bar z,t)$ is a generating function of complex solutions of (5.1)
where
$$P(z,\bar z,t) = -\int_0^{\bar z} a\,d\bar z + z^{\frac{1}{2}}\Big[-4\int_0^{\bar z} F\,d\bar z\Big]^{\frac{1}{2}} t.$$

We can now apply our theorems to some of these equations. Consider, for example, the case where

(i) $a(z,\bar z) = z\sum_0^{\sigma} c_n(z\bar z)^n$, c_n real. Hence a_z is also real.

(ii) $c(z,\bar z) = a_z + |a|^2 - \gamma_0^2$, γ_0 real. Hence c is also real.

Then $a_z = \bar a_{\bar z}$, $F = \gamma_0^2$ and $F_z = 0$ whence
$$E(z,\bar z,t) = \exp P(z,\bar z,t) = \exp\Big[-\sum_0^{\sigma}(c_n/n+1)(z\bar z)^{n+1} + 2(z\bar z)^{\frac{1}{2}}\gamma_0 t\Big]$$
$$E^{(1)}(z,s,t) = \exp\Big[-\sum_0^{\sigma}(c_n/n+1)s^{2(n+1)} + 2s\gamma_0 t\Big]$$
and hence $E^{(1)}(z,s,t) \geqq 0$.

We may now consider the case where

(iii) $a(z,\bar z) \neq 0$, $c_\sigma < 0$

then

$E^{(1)}(z,s,t)$ has the magnitude of $\exp[-(c_\sigma/\sigma+1)s^{2(\sigma+1)}]$

$p_n(s) \geqq c\exp[-(\mu c_\sigma/(\sigma+1))s^{2(\sigma+1)}]$ for $0<\mu<1$, μ fixed.

Applying Theorem 4, we have, if $f(z) = a_0 + a_k z^k + \cdots$ $a_0 \neq 0$, $a_k \neq 0$,

[8] This notation has the advantage that quantities like $u = z^r\bar z^s$ may be differentiated formally i. e. $u_z = rz^{r-1}\bar z^s$, $u_{\bar z} = sz^r\bar z^{s-1}$.

$$(5.2)\quad (q-1)m[s,u(z,\bar z)] \leq [\log s/\tilde r][\sum_1^q n[s,(u-\alpha_\nu)^{-1}]] + 8\log^+\log^+ h(s) + (\epsilon + |qc_\sigma|/(\sigma+1)s^{2(\sigma+1)}$$

for s large enough and such that $u(z,\bar z) \neq \alpha_\nu$ on $|z| = s$. Applying Theorem 5, we have, if $f(z) = a_\lambda z^\lambda + \cdots \quad a_\lambda \neq 0,\ \lambda > 0$. Then $u(z,\bar z)$ assumes all finite values.

An example of one of the equations considered above is

$$u_{z\bar z} + (-z)u_{\bar z} + (-\bar z)u_{\bar z} + [-1 + |z|^2 - 1]u = 0$$

i. e.

$$(\Delta u/4) - xu_x - yu_y + (x^2 + y^2 - 2)u = 0$$

and

$$E^{(1)}(z,s,t) = \exp[s^2 + 2st].$$

For some of the equations above we can get other classes of solutions. If the following conditions are satisfied

(a) $F_z = 0 \qquad F = -a_z - |a|^2 + c(z,\bar z)$

(b) $F = (a_z - \bar a_z/2)$

then we have for E (See [2, § 3])

$E(z,\bar z,t) = \exp[P(z,\bar z,t)]$ where

$P(z,\bar z,t) = -\int_0^{\bar z} a\,d\bar z + z[c + 2\int_0^{\bar z} F\,d\bar z]t^2$, c an arbitrary constant.

It can be shown without much trouble that our theorems apply to these new classes of solution if

$$a(z,\bar z) = z\sum_{n=0}^{\sigma} c_n(z\bar z)^n,\ c_n \text{ real},\ c(z,\bar z) = a_z + |a|^2.$$

6. Pseudo-Simple Functions. A classical theorem in the theory of analytic functions of a complex variable states that if $f(z) = z + c_2z^2 + c_3z^3 + \cdots$ is simple [9] in the circle $|z| < 1$, then (See [6])

$$|c_m| \leq k(m) \qquad m = 2, 3, \cdots$$

where

$$k(2) = 2, \quad k(3) = 3, \quad k(4) = 4.2848, \quad k(5) = 5.9158 \text{ etc.}$$

There exists a conjecture that if $f(z)$ is simple in $|z| < 1$, $|c_m| \leq m$.

---

[9] A function $f(z)$ is said to be simple in a domain if $f(z)$ is analytic and simple valued and if for every two distinct points z_1 and z_2 of the domain $f(z_1) \neq f(z_2)$.

If $f(z) = a_0 + a_1 z + a_2 z^2 + \cdots$ is simple in a domain it can have no points of multiplicity greater than one. Thus if $f(z)$ is simple in $|z| < \rho$, then

$$(6.1) \qquad \tilde{n}[r,(f-\alpha)^{-1}] = 0 \text{ or } 1 \text{ for all } \alpha \text{ and all } r < \rho.$$

On the other hand if

$$(6.2) \qquad \tilde{n}[r,(f-\alpha)^{-1}] = 0 \text{ or } 1 \text{ for all } \alpha$$

then $f(z)$ is simple in $|z| < r$ and thus

$$(6.3) \qquad |a_m| \leq k(m) r^{m-1} |a_1| \qquad m = 2, 3, 4, \cdots.$$

We can now set up an analogue of the concept of simple functions and establish a theorem corresponding to the above for functions $u(z,\bar{z}) \varepsilon \mathcal{B}(\vec{E})$ where E is a generating function of the type we have been considering.

Definition. *Suppose that* $u(z,\bar{z}) = v_s(z)$ *for* $|z| = s$ *where* $v_s(z)$ *is an entire function of the two complex variables* z *and* s. *Then* $u(z,\bar{z})$ *is said to be pseudo-simple in* $|z| < \rho$ *if*

$$(6.4) \qquad \tilde{n}[s,(v_s(z)-\alpha)^{-1}] = 0 \textit{ or } 1 \textit{ for all } \alpha \textit{ and all } s \textit{ such that } 0 < s < \rho.$$

It is to be noted that if $u(z,\bar{z})$ has no α-points on $|z| = s$, then

$$n[s,(u(z,\bar{z})-\alpha)^{-1}] = \tilde{n}[s,(v_s(z)-\alpha)^{-1}] = 0 \text{ or } 1.$$

We have defined pseudo-simple in terms of $v_s(z)$ instead of $u(z,\bar{z})$ to avoid complications which occur when $u(z,\bar{z})$ has α-points on $|z| = s$. It would also have been possible to define this concept directly in terms of $u(z,\bar{z})$ and for a more general set of functions $u(z,\bar{z})$.

If the function $u(z,\bar{z})$ is pseudo-simple in $|z| < \rho$, it does not necessarily follow that the inverse function of $u(z,\bar{z})$ is single valued as is the case for simple functions. For example consider the case of

$$u(z,\bar{z}) = (zJ_1(|z|)/|z|) = (z/2)[1-(1/8)(z\bar{z}) + \cdots]$$

Then $v_s(z) = zJ_1(s)/s$. If $J_1(s) \neq 0$, $\tilde{n}[s,(v_s(z)-\alpha)^{-1}] = 0$ or 1 for each α. If s_1 is the smallest $s > 0$ such that $J_1(s) = 0$ then our function is pseudo simple in $|z| < s_1$.

On the other hand if ϵ is a positive number which is small enough, then $J_1(s)$ assumes the value ϵ twice in the interval $0 < s < s_1$. Hence $u(z,\bar{z})$ assumes ϵ twice in $|z| < s_1$. The reason that $\tilde{n}$ remains less than two is that the index of the second ϵ point is -1 and cancels that of the first.

From our definition we see that if $u(z,\bar{z})$ is pseudo-simple in $|z|<\rho$ and $v_s(z)=\Sigma c_{mn}z^m s^n$ then

$$(6.5) \qquad |c_m(s)| \leq k(m)s^{m-1}|c_1(s)|, \qquad m=2,3,\cdots, \quad 0<s<\rho$$

where
$$c_m(s)=\sum_{n=0}^{\infty} c_{mn}s^n.$$

If $u(z,\bar{z})=\Sigma A_{mn}z^m\bar{z}^n$, $m>n$, then $v_s(z)=\Sigma A_{mn}z^{m-n}s^{2n}$. Thus we have $c_m(s)=\Sigma A_{m+n,n}s^{2n}$. Furthermore, as was pointed out in (2.4), if $u(z,\bar{z})\,\varepsilon\,\mathscr{B}(E)$ where $E^{(1)}(z,s,t)=\Sigma E_\mu^{(m)}(t)z^\mu s^m$ then

$$c_{mn}=\sum_{\nu=0}^{n}\frac{A_{\nu 0}\Gamma(v+1)}{\pi^{1/2}\Gamma(v+\frac{1}{2})}\int_{-1}^{1}E_{n-\nu}^{(m)}(t)(1-t^2)^{\nu-\frac{1}{2}}dt.$$

Thus we have

THEOREM 6. *If $u(z,\bar{z})=\Sigma A_{mn}z^m\bar{z}^n$ belongs to $\mathscr{B}(E)$ where $E^{(1)}(z,s,t)$ $\Sigma E_\mu^{(m)}(t)z^\mu s^m$ and $u(z,\bar{z})$ is pseudo-simple in $|z|<\rho$, then*

$$(6.5) \qquad |c_m(s)| \leq k(m)s^{m-1}|c_1(s)| \qquad m=2,3,\cdots \quad 0<s<\rho$$

where

$$(6.6) \qquad c_m(s)=\sum_{n=0}^{\infty}A_{m+n,n}s^{2n}$$

and also

$$(6.7) \qquad c_m(s)=\sum_{n=0}^{\infty}\sum_{\nu=0}^{n}\frac{A_{\nu 0}\Gamma(v+1)}{\pi^{1/2}\Gamma(v+\frac{1}{2})}\left[\int_{-1}^{1}E_{n-\nu}^{(m)}(t)(1-t^2)^{\nu-\frac{1}{2}}dt\right]s^n.$$

Theorem 6 establishes as necessary conditions for $u(z,\bar{z})\,\varepsilon\,\mathscr{B}(E)$ to be pseudo-simple certain relations expressed as inequalities in terms of the coefficients of the expansion of $u(z,\bar{z})$. When (6.6) is used we consider for each m an inequality in terms of the coefficients $A_{m+n,n}$. When (6.7) is used we obtain for each m an inequality in terms of $\{A_{\nu 0}\}$. In these inequalities the functions $E_m{}^n(t)$ occur. These functions are independent of the particular function $u(z,\bar{z})\,\varepsilon\,\mathscr{B}(E)$ which we are considering. It is noteworthy that instead of (6.7) we can get relations involving $\{A_{\nu r}\}$ for any fixed r.

In the case where $E(z,\bar{z},t)$ is a generating function of the first kind i.e. $E(z,0,t)=c_1, E(0,\bar{z},t)=c_2$ (See [1]), it is possible to relax the condition on $u(z,\bar{z})$. It is sufficient to assume that $u(z,\bar{z})$ is an analytic function of the real variables x and y everywhere in the finite plane. This is enough because this condition implies that the associate function $f(\zeta)$ is entire (See [3, p. 300]) and then $u(z,\bar{z})$ can be expanded in the power series $\Sigma A_{mn}z^m\bar{z}^n$, $m>n$.

Appendix. A Modification of Nevanlinna's Theorem. The second theorem of Nevanlinna was stated in a form which was more convenient for our purpose than the one presented in [4, § 33 and 34]. However our modification can be established by a simple change of Nevanlinna's arguments. We start from ([4], p. 65, (15))

$$(7.1) \qquad m[r, F] > \sum_{\nu=1}^{q} m[r, (f-\alpha_\nu)^{-1}] - q \log(2q/\delta) - \log 3$$

$$F = \sum_{\nu=1}^{q} (1/f - \alpha_\nu)$$

and from [4, p. 66]

$$(7.2) \qquad m[r, F] \leq 2T(r, f) + m(r, (f'/f)) + m(r, \Sigma(f'/f - \alpha_\nu)) - m(r, f) - N_1(r) + \log |(1/kc_k)|$$

where $f(z) = c_0 + c_k z^k + \cdots$ is a meromorphic function and $c_0 \neq 0$, $c_k \neq 0$ and

$$N_1(r) = [2N(r, f) - N(r, f')] + N(r, 1/f').$$

Now, if we asume that $f(z)$ is entire, then

$$N(r, f) = N(r, f') = 0$$

$$N_1(r) = N(r, 1/f') \geq 0 \text{ for } r \geq 1$$

by the definition of $N(r, f)$ given in [4, p. 6].

We also have $m[r, f] = T[r, f]$. Hence (7.1) and (7.2) together give

$$(7.3) \qquad \sum_{\nu=1}^{q} m[r, (f-\alpha_\nu)^{-1}] \leq q \log(2q/\delta) + \log 3 + m(r, f) + m(r, f'/f) + m(r, \Sigma(f'/f - \alpha_\nu)) - N_1(r) + \log |(1/kc_k)|$$

or

$$(7.3') \qquad \sum_{\nu=1}^{q} m[r, (f-\alpha_\nu)^{-1}] \leq m(r, f) + S_0(r) - N_1(r)$$

where

$$(7.4) \qquad S_0(r) = m(r, \Sigma(f'/f - \alpha_\nu) + m(r, f'/f) + \log |3/kc_k| + q \log 2q/\delta.$$

From [4, p. 67] we have

$$(7.5) \qquad m(r, \Sigma(f'/f - \alpha_\nu)) < 24 + 4 \log 3 + 4 \log q + 4 \log^+ \log^+ R + 2 \log^+ 1/r + 4 \log^+ \rho + 3 \log^+ 1/\rho - r + 4 \log^+ T(\rho, f) + 3 \log^+ |1/\Phi(0)|$$

where $\log^+ |1/\Phi(0)| \leq q \log^+ R_0$.

In addition we have from [4, p. 61]

$$(7.6)\qquad m(r,f'/f) < 24 + 3\log^+|1/c_0| + 2\log^+ 1/r + 4\log^+\rho + 3\log^+ 1/\rho - r + 4\log^+ T(\rho, f)$$

and because $R_0 \geq |1/c_0|$, we have

$$(7.7)\qquad S_0(r) < 56 + \log^+|1/kc_k| + 6q\log^+ R_0 + 4\log^+\log^+ R + 2q\log 2q/\delta + 4\log^+ 1/r + 8\log^+\rho + 6\log^+ 1/\rho - r + 8\log^+ T(\rho, f).$$

From [4, p. 11] we have,

$$(7.8)\qquad T[r, (f-a)] - T[r, 1/f-a] = \log|c'_\lambda|$$

if $f(z) - a = c'_\lambda z^\lambda + c_{\lambda+1}z^{\lambda+1} + \cdots$ and $f(z)$ is non-constant.

This together with

$$(7.9)\qquad |T[r,f] - T[r,f-a]| \leq \log^+|a| + \log 2$$

gives

$$(7.10)\qquad m[r,(f-a)^{-1}] + N[r,(f-a)^{-1}] + \log|c'_\lambda| + \log^+|a| + \log 2 \geq m[r,f]$$

whence (7.3′) becomes

$$(7.11)\qquad (q-1)m[r,f] \leq \sum N[r,(f-\alpha_\nu)^{-1}] - N_1(r) + S_1(r)$$

where

$$(7.12)\qquad S_1(r) \leq S_0(r) + \sum\log^+|\alpha_\nu| + q\log 2 + \sum\log|c_0 - \alpha_\nu| \quad \text{if } |c_0 - \alpha_\nu| \neq 0$$

and we obtain our inequality for $S_1(r)$.

$$(3.2)\qquad S_1(r) < 56 + \log^+|1/kc_k| + 6q\log^+ R_0 + (4+q)\log^+ R + 2q\log 2q/\delta + 4\log^+ 1/r + 8\log^+\rho + 6\log^+ 1/\rho - r + 8\log^+ T(\rho, f) + \sum\log|c_0 - \alpha_\nu| + q\log 2.$$

Now if $f(z) = \sum_{\lambda}^{\infty} c_n z^n$ is entire where $\lambda > 0$, these arguments must be revised. In the main we shall make use of the same attack.

Equation 7.1 still holds. However the proof of 7.2 in Nevanlinna involved c_0. We can go through a similar argument with very few changes. From [4, p. 65] we have

$$(7.13)\qquad m[r,F] \leq m[r,1/f] + m[r,f/f'] + m[r,\Sigma f'/(f-\alpha_\nu)]$$

and (7.8) applied when $a=0$ gives

$$(7.14)\qquad m(r,1/f) = T(r,f) - N[r,1/f] + \log|1/c_\lambda|.$$

We can apply (7.3) to the function $g = f'/f = \lambda/z + \cdots$ and

$$(7.15)\qquad m(r,f/f') = N[r,f'/f] - N[r,f/f'] + \log|1/\lambda| + m[r,f'/f]$$

$$(7.16)\qquad m[r,F] \leq 2T(r,f) + m[r,\Sigma f'/(f-\alpha_\nu)] - m(r,f) - N_1(r) + \log|1/\lambda c_\lambda| + m[r,f'/f]$$

where $N_1(r) = N(r,1/f) + N(r,f/f') - N(r,f'/f) \geq 0$ for $r \geq 1$.

Combining this with (7.1) we have equation (7.3′) where

$$(7.17)\qquad S_0(r) = m[r,\Sigma f'/(f-\alpha_\nu)] + m(r,f'/f) + \log||3/\lambda c_\lambda| + q\log 2q/\delta.$$

Now we may make use of the footnote in [4, p. 61].

$$(7.18)\qquad m(r,f'/f) < 34 + 5\log^+|\lambda| + 3\log^+|1/c_\lambda| + 7\log^+ 1/r + 4\log^+\rho + 3\log^+(1/\rho - r) + 4\log^+ T(\rho,f).$$

The inequality (7.18) was proved on the assumption that the expansion of f has at least two non-vanishing terms. However if $f(z) = c_\lambda z^\lambda$, $f'/f = \lambda/z$ and thus

$$m[r,f'/f] = \log^+|\lambda/r| \leq \log^+|\lambda| + \log^+ 1/r$$

and the inequality (7.18) holds.

Now, on the assumption that $c_0 \neq \alpha_\nu$, i.e. $\alpha_\nu \neq 0$, inequality (7.5) is derived where

$$R_0 = \max\{|1/c_\lambda|, |1/\alpha_\nu|\}$$

and

$$\Phi(0) = \Pi(c_0 - \alpha_\nu) = \Pi(-\alpha_\nu)$$

and thus

$$\log^+|1/\Phi(0)| \leq q\log^+ R_0 \text{ as before.}$$

Combining (7.5), (7.18), (7.17), we have, for $q \geq 1$, and on making a few simplifications

$$(7.19)\qquad S_0(r) < 66 + \log^+|1/\lambda c_\lambda| + 6q\log^+ R_0 + 4\log^+\log^+ R + 2q\log 2q/\delta + 9\log^+ 1/r + 8\log^+\rho + 6\log^+(1/\rho - r) + 8\log^+ m(\rho,f) + 5\log^+|\lambda|.$$

Now from equation (7.3′) we can go to (7.11) by making use of (7.10). In (7.10) c'_λ stands for the first coefficient of $f(z) - a$. If $a \neq 0$, we have $c'_\lambda = -a$ because there is no constant term in $f(z)$. Thus we have

$$(7.11) \qquad (q-1)\; m[r,f] \leq \sum N[r,(f-\alpha_\nu)^{-1}] - N_1(r) + S_1(r)$$

where

$$(7.12) \qquad \begin{aligned} S_1(r) &\leq S_0(r) + \sum \log^+ |\alpha_\nu| + q \log 2 + \sum \log |\alpha_\nu| \\ &\leq S_0(r) + 2q \log^+ R + q \log 2 \end{aligned}$$

and thus

$$(3.2a) \qquad \begin{aligned} S_1(r) < 66 &+ \log^+ |1/\lambda c_\lambda| + 6q \log^+ R_0 + (4+2q) \log^+ R + \\ &+ 2q \log(2q^2/\delta) + 9 \log^+ (1/r) + 8 \log^+ \rho + 6 \log^+ (1/\rho - r) \\ &+ 8 \log^+ m(\rho, f) + 5 \log^+ |\lambda| + q \log 2. \end{aligned}$$

Finally, we have, if $\tilde{n}[r_1, (f-\alpha_\nu)^{-1}] = 0$, $r > r_1$

$$N[r, (f-\alpha_\nu)^{-1}] = \int_{r_1}^{r} (\tilde{n}[t,(f-\alpha_\nu)^{-1}]/t)\, dt \leq \tilde{n}[r,(f-\alpha_\nu)^{-1}] \log r/r_1$$

whence we obtain our desired result

$$(3.1) \qquad \begin{aligned} (q-1) m[r,f(z)] \leq [\log r/r_1] [\sum_{\nu=1}^{q} \tilde{n}[r,(f-\alpha_\nu)^{-1}]] \\ - N_1(r) + S_1(r). \end{aligned}$$

BROWN UNIVERSITY.

BIBLIOGRAPHY.

1. Stefan Bergman, "Linear operators in the theory of partial differential equations," *Trans. Amer. Math. Soc.*, vol. 53 (1943), pp. 130-155.
2. Stefan Bergman, "The determination of some properties of a function satisfying a partial differential equation from its series development," *Bull. Amer. Math. Soc.*, vol. 50 (1944), pp. 535-546.
3. Stefan Bergman, "Certain classes of analytic functions of two real variables and their properties," *Trans. Amer. Math. Soc.*, vol. 53 (1945), pp. 299-331.
4. Rolf Nevanlinna, "*Le théorème De Picard-Borel et la théorie des Fonctions méromorphes.*"
5. Alexandroff-Hopf: *Topologie* I.
6. Alfred Rosenblatt, "Nouvelles recherches sur les coefficients des séries univalentes," *Revista de Ciencias* (1938).
7. Nielsen and Ramsay, "On particular solutions of linear partial differential equations," *Bull. Amer. Math. Soc.*, vol. 49 (1943), pp. 156-162.

THE INTRODUCTION OF LOCAL CONNECTIVITY BY CHANGE OF TOPOLOGY.*

By Gail S. Young, Jr.

1. Introduction. Suppose that S is a topological space, and that G is a collection of subsets of S. I shall say that a point P of S is a G-limit point of a subset M of S provided that every open neighborhood of P contains an element of G that intersects both P and $M - P$. A set is G-closed if it contains all its G-limit points, and is G-open if its complement is G-closed. These definitions determine a new topology for S—the G-topology—which is actually a "topology" in the sense that if S is a T_0 space originally it is still one in the G-topology. This is shown in Theorem 1. This G-topology is consistent with the original topology in that every G-limit point of a set is a limit point of that set in the original topology, but is weaker [1] in the sense that there are "less" limit points. The motivation of this paper is in the fact that for some choices of G, S is locally connected in the G-topology, though it may not have been so originally. This is the case, for example, if G is the collection of all arcs of S. When this does occur, it is often possible to prove a theorem about S by changing to a G-topology and using properties of locally connected spaces. For example, consider the following theorem, due to Whyburn: [2] A complete metric space S is arcwise connected if and only if each two points of S can be joined by a connected and locally connected subset of S. By taking G to be the collection of all locally connected and connected subsets of S, we find from Theorems 3, 4, and 5 that in the G-topology S is still complete metric and connected, and is now locally connected. Arcwise connectivity follows immediately. In a similar fashion, virtually any theorem on general locally connected spaces can be used to give some sort of theorem on non-locally connected spaces—though usually a theorem of no particular interest.

Rather little is yet known about general locally connected spaces.[3] It seems likely that the concepts of this paper will become more useful when

* Received January 4, 1946; presented to the American Mathematical Society, September 17, 1945.

[1] Cf. Birkhoff [1]. Numbers in square brackets refer to the bibliography.

[2] Essentially the second theorem on page 756 of [23].

[3] Of items occurring in the bibliography of this paper, [8], [14], [21], and [22] are particularly important in this connection.

more results are available. The notion of a G-topology may possibly be useful when the resulting space is not locally connected, but I have found no example of this.

One application of these concepts seems interesting enough to deserve special mention. That is the following theorem: *Let M be any set that contains no arc, let D be a dendrite, and let H be a collection of arcs such that each point of M is joined to D by some arc of H, and such that $M + D + H^*$ contains only one arc between any two points of M. Then $M + D + H^*$ has the fixed-point property.*[4] Section **4** is devoted to the proof of this result. As a preliminary, study is made in Section **3** of a type of generalized dendrite, with particular emphasis on the fixed-point property.

There is a strong relationship between the concepts of this paper and certain work of Menger, Whyburn, and Myers. This will be discussed at the end of Section 2.

2. G-Topologies. Notation: We shall often use a description of the type of sets forming the collections G as a prefix, in place of the letter G. Thus if G is the collection of all arcs of S, we shall speak of the arc-topology of S, or the a-topology. Also for the cases where G is the collection of (1) all connected subsets of S; (2) all locally connected subsets; or (3) all rectifiable arcs, we shall use the abbreviations (1) c-topology; (2) lc-topology; or (3) r-topology.[5] The original topology of S will be referred to as the o-topology. To avoid confusion, names of topological properties will be similarly prefixed to indicate in which topology a set has the property: thus, a set may be o-connected, lc-compact, c-open, etc.

Examples: (1) If S is the closure of the graph of $y = \sin 1/x$, then the a-, c, lc-, and r-topologies are equivalent, and in any of these S is the sum of three components, two being open curves, and the third an arc.

(2) If S is the Cantor star—the join of a Cantor set and a point—then in any of the four topologies used in (1) S is, roughly speaking, the sum of uncountably many intervals all mutually perpendicular.

(3) If S is one of the locally connected spaces due to Moore [13] or to Kuratowski and Knaster [10] that contains no arc, then the c- and lc-topologies are equivalent, while the a-topology is discrete. In Moore's example,

[4] If H is a collection of sets, H^* denotes the sum of the elements of H. A set K has the fixed-point property provided that for every continuous transformation of K into itself, some point is its own image.

[5] This notation was introduced by Wallace [19].

which is compactly connected, the compact-continuum topology is equivalent to the lc-topology, while in the other one, which contains no perfect set, the compact-continuum topology is discrete.

THEOREM 1. *If S is a T_0 space in its o-topology, then it is a T_0 space in any G-topology, and every o-open set is G-open. If every element of G is o-connected, then every o-component of an o-open set is G-open.*

Proof. Clearly, S and the empty set are both G-open. Let D be the sum of a collection of G-open sets, and let P be a point of one of them, say E. The statement that E is G-open implies that there is an o-open set U containing P such that no element of G lies in U and intersects both P and $S - E$. Then certainly no element of G lies in U and intersects both P and $S - D$, so that D is G-open.

The product of two G-open sets is G-open, for let D and E be two G-open sets and let P be a point of their intersection. There exist o-open sets, U and V, containing P, such that no element of G lying in U (or V) intersects P and $S - D(S - E)$. The set $U \cdot V$ is o-open and contains no element of G that intersects $S - D \cdot E$. Hence P is not a G-limit point of $S - D \cdot E$, and it follows that $D \cdot E$ is G-open. The rest of the theorem is easily proved.

THEOREM 2. *If S is a Hausdorff space in the o-topology, it is also a Hausdorff space in the G-topology.*

THEOREM 3. *If every element of G is G-connected, and S is a T_0 space, then S is G-locally connected, and if, in addition, every two points of S lie in an element of G, then S is G-connected. Hence S is a- and lc-locally connected.*

Proof. Let H be a G-component of a G-open set, D. If H is not open, it contains a point P which is a G-limit point of $D - H$. Since each element of G is G-connected, no element of G that contains P and intersects $D - H$ lies entirely in D, so that each such element intersects the G-boundary of D. But then every o-neighborhood of P contains an element of G that intersects P and the G-boundary of D, so that P is in the G-boundary of D, which is impossible. It follows that S is G-locally connected. The rest of the theorem is obvious.

It may be thought that for any collection G, of connected sets, S would be G-locally connected, but this is not the case. An example which shows this will be given following Theorem 6.

A question of considerable importance is this: Suppose that S has a strong

topology, say compact metric. By Theorem 2, for any collection G, S is Hausdorff in the G-topology. Even if S were G-locally connected, however, in general one would prefer to keep the power of compact metricity and not have the local connectivity with the weak separation axioms. Will S always have a topology that is not much weaker than its original one? For example, will S always be metric if it is originally compact metric? Unless extra conditions are imposed on G, such as those of Theorem 4, the answer to this is no. As a simple example, let S be the plane, and let G be the collection of all arcs having no more than one point in common with the x-axis. Then S is certainly not metric, nor even a Moore space in the G-topology, for if P is a point of the x-axis, the G-closure of every G-open set that contains P intersects the complement of P in the x-axis, which is G-closed.

THEOREM 4. *Let G have the property that if A is a point of an o-domain, D, there exist two o-open sets, R and R', such that* (1) *R contains R' and R' contains A, and* (2) *if C is a simple chain*[6] *of elements of G, one link of which contains A and all of whose links lie in R', and X is any point of C^*, then there is an element of G which lies in R and contains $X + A$. Then if S is metric or a Moore space*[7] *in the o-topology, it is still metric or still a Moore space in the G-topology.*

Proof. Suppose first that S is a Moore space; let $G_1, G_2, G_3, \cdots$ be the collection of o-regions postulated by Axiom 1 of Moore's book [14]. For each natural number n, let H_n be the collection of all G-open sets R defined as follows: If P is a point of an o-region U of G_n, let R be either the maximal subset of U containing P such that every two points of R can be joined by a simple chain of elements of G lying in R; or if this set is empty, let R be P. It is easy to see that R is always G-open. Then the collection $H_1, H_2, H_3, \cdots$ satisfies the first two parts of Axiom 1. Let E be a G-open set, and let A and B be points of E. To show that S is a Moore space in the G-topology, we need only show that there is a natural number n such that if h is an element of H_n that contains A, the G-closure of h is a subset of $E - B$. There is a natural number n_1 such that if R is an o-region of G_{n_1} that contains A then (1) the o-closure of R does not contain B, and (2) no element of G that

---

[6] A collection $H_1, H_2, \ldots, H_n$ of sets is a simple chain from the point A to the point B provided that any two successive sets have a point in common; sets that are not successive do not intersect; H_1 is the only set that contains A; and H_n is the only set that contains B. The sets H_i are the links of the chain. Cf. Moore [14], p. 56.

[7] A Moore space is a space satisfying the first three parts of Axiom 1 of Moore's book [14].

contains A and lies in R intersects $S - E$. Let D be some definite o-region of G_{n_1} that contains A, and choose two o-domains R and R' containing A and satisfying for D and A the conditions of the hypothesis. There is an integer n_2 such that the o-closure of every G-region of H_{n_2} that contains A lies in R'. Suppose that for some integer n greater than n_2 there is a G-region, F, of H_n that contains A, and whose G-closure intersects $(S - E) + B$. Then there is an element g of G that contains a point, A', of F, intersects $(S - E) + B$, and lies in R'. By the definition of $H_1, H_2, H_3, \cdots$, there is a finite chain, C, of elements of G that lies in F, and hence in K', and contains both A and A'. By our hypothesis there is, then, an element g' of G that lies in R and contains A and a point of $[(S - E) + B] \cdot g$. But this contradicts the definition of n_1. It follows that if the topology defined by the collections $H_1, H_2, H_3, \cdots$ is the G-topology, then S is a Moore space in the G-topology. But clearly our hypothesis and the definition of these collections show that this is so.

Suppose now that S is metric. We shall define a G-metric. Suppose that no distance in S is greater than 1.[8] For each two points A and B of S define the G-distance $\rho(A, B)$ as follows: (1) If there is a simple chain of elements of G from A to B, let $\rho(A, B)$ be either 1 or the greatest lower bound of the sum of the diameters of the links of such chains, whichever is the lesser; (2) if there is no such chain, let $\rho(A, B) = 1$. Let $\rho(A, A) = 0$. We need only show that this metric satisfies the triangle inequality and that it is equivalent to the G-topology. Let A, B, and C be three points. If $\rho(A, B) + \rho(B, C)$ is less than $\rho(A, C)$, there exist simple chains, H and K, of elements of G from A to B and from B to C, respectively, such that the sum of the o-diameters of the links of H and K is less than $\rho(A, C)$. But $H + K$ contains a simple chain, L, of elements of G from A to C, and the sum of the o-diameters of the links of L is less than $\rho(A, C)$, which is impossible.

Clearly, convergence in the G-topology implies convergence in the ρ metric. Suppose that β is a sequence of points converging to a point, X, in the ρ metric. Let ϵ be any positive number. There exist two o-open sets R and R' which contain X and are of o-diameter less than ϵ, and which satisfy with respect to S and to X the conditions of the hypothesis. Let Y be any point of β such that $\rho(X, Y)$ is less than half the o-distance, d, from X to $S - R$. Then there is a simple chain, H, of elements of G from X to Y such that the o-diameter of H^* is less than d. It follows by hypothesis that there is an element of G lying in R and containing X and Y. Hence β

[8] This involves no loss of generality; any metric can be made to satisfy this by changing distances greater than 1 to 1.

G-converges to X, which completes the proof that the ρ metric and the G-topology are equivalent.

THEOREM 4(a). *Any one of the following conditions implies the condition of the hypothesis of Theorem 4: (1) the sum of two intersecting elements of G is an element of G; (2) each two points of the sum of two intersecting elements of G are joined by an element of G lying in that sum; and (3) if D is an o-domain, each two points of the sum of two intersecting elements of G that lie in D can be joined by an element of G lying in D. In particular, the collections of all arcs, all connected subsets and of all connected and locally connected subsets satisfy that hypothesis.*

The G-distance function, $\rho(A,B)$, cannot in general be replaced by the function $\rho'(A,B)$ defined as the greatest lower bound of diameters of elements of G containing A and B, or as 1, if none exists. We do have the following result.

THEOREM 4(b). *If S is semi-metric, for any collection G of subsets of S, the $\rho'(A,B)$ function defines a semi-metric agreeing with the G-topology of S.*

In general, even for a collection satisfying any of the conditions of Theorem 4(a) it is not true that completeness in the o-metric will imply completeness in any G-metric. A simple example of this is to let S be the plane and to let G be the collection consisting of all points of the x-axis with irrational abscissae, and of all connected sets containing no such point. The following theorem covers the most important cases.[9]

THEOREM 5. *Suppose that S is metric in its o- and G-topologies, and is o-complete. Suppose further that G is such that if $g_1, g_2, g_3, \cdots$ is a sequence of elements of G o-converging to a point P, and such that $g_1 + g_2 + \cdots$ is connected, then for every o-open set R containing P there is an integer N such that, for every $n > N$, R contains an element of G intersecting both P and g_n. Then in the G-metric, $\rho(A,B)$, S is complete.*

Proof. Suppose that $P_1, P_2, P_3, \cdots$ is a subsequence of a Cauchy sequence in the G-metric. Then this subsequence o-converges to a point P. For each integer j there is an integer N_j such that if n and m are larger than

[9] Considerable difficulty arises in the problem of G-completeness for an o-complete Moore space, due to the possible presence in G_1, G_2, $G_3, \ldots$ of "large" regions which really have nothing to do with the topology of the space. All I can say is that the collections of G-regions defined in Theorem 4 will not always satisfy part (4) of Axiom 1, though this, of course, does not imply that no set of collections will.

N_j, P_n and P_m are each at o-distance less than $1/2^j$ from P and are at G-distance less than $1/2^j$ from one another. It follows that for each two such points there is a simple chain of elements of G from P_n to P_m such that the sum of the o-diameters of its links is less than $1/2^j$; hence the sum of the chain is at o-distance less than $2/2^j$ from P. It is easy to select a subsequence $Q_1, Q_2, Q_3, \cdots$ of $P_1, P_2, P_3, \cdots$ such that for each n Q_n is at o-distance less than $1/2^n$ from P and such that there is a finite chain G_n of elements of G joining Q_n and Q_{n+1}, the sum of their o-diameters being less than $1/2^n$. The sequence of all elements of G which are links of any one of $G_1, G_2, G_3, \cdots$ clearly o-converges to P. It follows from the hypothesis and the definition of the G-topology that $Q_1, Q_2, Q_3, \cdots$ G-converges to P. Since, then, every subsequence of a G-Cauchy sequence, β, contains a G-convergent subsequence, it follows that β G-converges.

THEOREM 5(a). *If S is o-metric and o-complete, it is complete metric in each of the a-, lc-, and c-topologies.*

Proof. To take one case, if a sequence of arcs o-converges to a point P and has an o-connected sum, then this sum plus P is an o-continuous curve, and the condition of Theorem 5 is easily seen to hold.

Of particular interest for application is the question of the behavior of transformations of S under changes of topology. Of course, the continuity of a transformation of S depends not only on the topology chosen for S, but also on the topology chosen for the image of S. We shall consider, however, only one of the many possible theorems.

THEOREM 6. *If a transformation $T(S) = U$ of a T_0 space S onto a space U is o-continuous, it is G-continuous for any collection G. If further, H is a collection of subsets of U such that for any element g of G each two points of $T(g)$ are joined by an element of H lying in $T(g)$, then T is G-continuous in the H-topology of U.*

Proof. The first statement is obvious. We prove the second. Suppose that K is a subset of S and that X is a point of the G-closure of K. Since T is o-continuous, $T(X)$ is contained in the o-closure of $T(K)$ (that is, the closure of $T(K)$ in the original topology of U). If it is not a point of $T(K)$, every o-open subset R of U that contains $T(X)$ also contains the image of some element g of G that intersects X and K. By hypothesis, R then contains an element of H that intersects both $T(X)$ and $T(K)$. It follows that $T(X)$ is in the H-closure of $T(K)$, which completes the proof.

Except to give the following example, of interest in connection with the

work of Whyburn discussed below, I shall not discuss the equivalence and comparison of various types of G-topologies. This example shows that the c- and lc-topologies are not equivalent.

Example: Let T denote the plane rectangle with the points $(0, \pm 1)$ and $(1, \pm 1)$ as vertices; let R denote the closure of the set of points of the graph of $y = \sin \pi/x$ for which $0 < x < 1$; let P denote the point $(1/2, 0)$ and Q denote the point $(1/3, 0)$. For each positive integer n, let R_n, P_n, and Q_n denote the images of R, P, and Q, respectively, under a sense-preserving affine transformation of the plane throwing T into the rectangle whose vertices are $(1/(n+1), \pm 1/n)$, $(1/n, \pm 1/n)$; and let A_n denote the arc joining Q_n to P_{n+1} consisting of two vertical intervals and an interval of the line $y = 1$. Let I denote the interval $(0,0)$, $(0,1)$ of the y-axis, and let S denote $I + A_1 + R_1 + A_2 + R_2 + \cdots$, topologized by the plane metric. Then in the lc- or a-topologies the sets I and $S - I$ are mutually separated, while in the c-topology S is still connected.

The viewpoint of this section has some precedence in the literature. Mazurkiewicz [11] defined and studied for compact locally connected spaces, K, a metric which replaced the distance between each two points of K by the greatest lower bound of diameters of continua containing the two points. Whyburn [24] (see also page 154 et seq. of his book [25]) extended this definition to general locally connected metric spaces by replacing the requirement that the sets defining the new metric be continua with the requirement that they be merely connected, pointing out that Mazurkiewicz' definition was inadequate if the spaces studied were not compactly connected. Whyburn exploited this metric—his relative distance transformation—primarily in studying plane domains.[10] From the standpoint of this paper, Mazurkiewicz was studying the compact-continuum topology and Whyburn the c-topology of their respective spaces, and using essentially the metric of Theorem 4. However, in each case, though the metric changed, the topology of the space did not. From another approach, ideas similar to those of this paper have been developed in metric geometry by Menger [12] and Myers [15]. For compact metric spaces M in which each two points are joined by an arc of finite length (and hence by a geodesic[11]) Menger defined a metric as the length of the geodesic joining each pair of points. In this geodesic metrization, M is convex, but not necessarily topologically the same. Myers

[10] Some of the general properties of his metric are valid in the more general spaces of this paper, but I have not explicitly stated these.

[11] For a discussion of these and related topics, see Chapter VI of Blumenthal's book [3].

extended this to rectifiably connected metric spaces in general by using for the new metric the greatest lower bound of lengths of arcs joining two points. He considered many of the properties of such spaces, proving results like many of those considered here, and using the properties of the new metric to obtain conditions for the existence of geodesics. His results are not special cases of those herein; if G is the collection of all rectifiable arcs, the G-metric is expressed in terms of the diameters and not of the lengths of these arcs, so that the G-metric can become small without the geodesic metric decreasing. A unified treatment is of course possible, but it seemed to me that the rather slight gain was not worth the extra space required. I should also mention the work of Hewitt [5] who in other connections used the same general idea of change of topology as a method of proof.

3. Generalized dendrites. In this section, I shall define a type of generalized dendrite and study its properties. I am particularly interested in deriving a certain condition for the fixed-point property to hold, but it will become clear that the methods used will permit a number of theorems involving mappings of dendrites to be proved for at least many generalized dendrites. It would be possible to shorten the length of this section by assuming that we are dealing with a metric space. Though this would not greatly affect the use I intend to make of the results of this section to illustrate the use of the preceding section, it seems preferable to obtain as much generality as possible.

Definition: By a generalized dendrite [12] is meant a locally connected Hausdorff space T such that if A and B are two points of T, and Δ_1 and Δ_2 are two simple chains of connected domains from A to B, and Δ_1 has more than two links, then some link of Δ_1 that does not contain A or B intersects some link of Δ_2. Throughout this section T denotes such a space.

THEOREM 7. *Under the hypotheses of the definition every link of Δ_1 intersects some link of Δ_2.*

Proof. Suppose that $\Delta_1 = D_1, D_2, D_3, \cdots$, where the numbering is in the order from A to B, and suppose that for some integer j, D_j intersects no link of Δ_2. Then there exist integers i and k such that $i < j < k$, and D_i and D_k intersect links of Δ_2, but no link of Δ_1 between D_i and D_k intersects a link of Δ_2. There exist two points, X and Y, in D_i and D_k, respectively such

[12] This is not the same as the generalized dendrite recently defined by Shanks [17], which is a generalization removing local connectivity. My generalization is along the line of Moore's dendron [14], Whyburn's denodular sets [22], and Wallace's trees [18]. In each of the spaces considered by these authors, their term is equivalent to mine.

that Δ_1 and Δ_2 contain subchains from X to Y. But clearly the existence of these subchains contradicts the definition of T.

THEOREM 8. *If the closed set H separates the point A from the point B in T then some point of H separates A from B in T.*

Proof. Let C_1 be the component of $T - H$ containing A and let C_2 be the component of $T - (\bar{C}_1 - C_1)$ that contains B. Suppose that the boundary of C_1 contains two points, X and Y. There exist mutually exclusive connected domains, D and E, containing X and Y, respectively. Then C_1, D, and C_2, and C_1, E, and C_2 are two chains contradicting the definition of T.

THEOREM 9. *If A and B are two points of T, they are the non-cut points of a bicompact and (sequentially) compact, locally connected, connected set, H, such that each point of $H - (A + B)$ separates A from B in T.*

Proof. Let H denote the set of all points P such that if Δ is a simple chain of connected domains from A to B, then Δ^* contains P. If the point X of $\bar{H} - (A + B)$ does not separate A from B, there is a simple chain, Δ, from A to B of connected domains no one of which contains X in its closure. Then $\overline{\Delta^*}$ contains H, but not X. Hence every point of $\bar{H} - (A + B)$ separates A from B, and therefore belongs to H. We now show that H is bicompact. Let G be any collection of domains covering H. Let G' denote the collection of components of $T - H$; there is a simple chain of connected domains from A to B every link of which is either a domain of G' or a component of a domain of G. Bicompactness follows readily.

If H is not connected, then there exist two points, X and Y, of H such that no point of H is between them in the separation ordering of H. There is a domain D containing X such that $\bar{D} \cdot Y = 0$. Then $\bar{D} - D$ separates X from Y in T. By Theorem 8, so does some point, Z, of $\bar{D} - D$. But then Z belongs to H, which is impossible.

The other properties of H follow from the results of F. B. Jones [7], since it is easy to see that H is linear in his sense, or they are not hard to prove directly.

THEOREM 9(a). *If T is separable, or is a Moore space, the set H of Theorem 9 is a true arc.*

Proof. The argument used in the proof of Theorem 6 of my paper [26] can be modified to prove that if L is a closed subset of a locally connected Hausdorff space, M, and K is a countable set of points of M dense in L such that each point of K either belongs to L or is separated from some point of

L by some other point of L, then L is separable. Hence H is separable, and, by Theorem 10 of Jones [7], is a true arc. For a Moore space, this is a consequence of Theorem 2, p. 432, of Moore [14].

THEOREM 10.[13] *Every connected subset of T is itself a generalized dendrite.*

Proof. If C is a connected, locally connected subset of T, clearly it follows from Theorem 9 that it is a generalized dendrite. Suppose that the connected subset C is not locally connected at a point A. There is a connected domain, R, of T containing A, such that $C \cdot R$ is not connected. From the argument in Theorem 9 each two points of R are joined by a "pseudo arc" lying in R. If its ends lie in C, so does the "pseudo arc." This gives a contradiction.

THEOREM 11. *If T is either* (1) *bicompact and separable; or* (2) *a separable Moore space having only a countable number of non-cut points, then T is metric.*

Proof. We shall show that T is metric under the first hypothesis. If K is a countable subset of T dense in T, select for each two points of K a point of T that separates them; let K' denote the set of all such points. Clearly K' is countable and dense in T. Let G be the collection of all connected open sets whose boundaries are finite subsets of K'; since there are only a countable number of finite subsets of K', it readily follows that G is countable. Let P be a point and R be a connected open subset of T containing P. For each point X of the boundary of R choose a point P_X of K' separating X from P; let D_X be the component of $T - P_X$ that contains X. Since $T - R$ is closed, a finite collection of domains D_X covers it. Let H denote the sum of their boundaries. Then H is a finite subset of K' and the component of $T - H$ that contains P lies together with its boundary in R. Consequently, T is regular in the sense of Alexandroff and Urysohn [25] and is completely separable; therefore it is metric.

In the second case, by a theorem of Jones [6], if every uncountable subset of T has a limit point, then T is metric. If K is an uncountable subset of T, the argument used in Theorem 72 of Chapter I of Moore [14] proves that some two points of T are separated by uncountably many points of K. It follows from Theorem 9(a) that K has a limit point.

Neither of the hypotheses of Theorems 9(a) or 11 can be weakened by the omission of any condition and still have the theorems remain true On

---

[13] Cf. Theorem 32, p. 116, of Moore [14].

the other hand the condition of separability alone lets us strengthen the topology of T into a metric space.

THEOREM 12. *If T is separable, then there is a biunivalued continuous transformation of T into a connected subset of a compact, metric dendrite. If in addition the sum of every monotone increasing sequence of arcs of T is contained in an arc of T, then this subset is closed.*

Proof. The argument given by Whyburn in [21] to prove a similar theorem for metric, separable potentially regular sets, and which is closely related to the proof of Theorem 11, is sufficient under our hypotheses to prove that T can be mapped in the desired way onto a metric generalized dendrite. The proof consists essentially in redefining "neighborhood" to mean a complementary domain of a finite subset of the set K' defined in Theorem 11, and proceeding as in that theorem to prove that T is metric in this new topology. Then either Theorem A_1 of Whyburn's [21] or Theorem 8.2 of his [22] shows that T can be mapped onto a subset M of a compact dendrite D by a transformation $f(T) = M$ of the desired type.

To prove the rest of the theorem, let X be a point of $\bar{M} - M$, if any exists. Then there is a sequence, $X_1, X_2, X_3, \cdots$ of points of M approaching X such that for each n X_n separates X_{n-1} from X in $\bar{M}$. The sequence X_1X_2, $X_1X_3, X_1X_4, \cdots$ of arcs of M is monotone increasing, and hence so is $f^{-1}(X_1X_2), f^{-1}(X_1X_3), \cdots$. In T there is an arc AB irreducibly containing all these arcs. Then $f(A + B) = X + X_1$, which proves that M is closed.

We are now in a position to prove the fixed-point theorem for generalized dendrites mentioned above. This theorem will be stated for a generalized dendrite which is a Hausdorff space, but the proof will be valid only for arcwise-connected generalized dendrites. The reason for this is that it is possible to give a proof for the general case merely by modifying the argument for Theorem 2.1 of Chapter XII of Whyburn's book [25], closed, connected subsets taking the place of the A-sets, and the existence of certain intersections being guaranteed by the condition of the theorem. But such a proof would have little novelty, and I prefer to give the more restricted one here, which indicates a general method for studying transformation properties of certain generalized dendrites by means of Theorem 12.

THEOREM 13.[14] *A sufficient condition that a generalized dendrite T have*

[14] Cf. Wallace [18], where he proves a similar theorem for certain point-to-set mappings of bicompact Hausdorff trees. My result generalizes his for point-for-point mappings. I have not obtained a full generalization.

the fixed-point property is that the sum of any increasing simple sequence of pseudo arcs of T be contained in a pseudo arc.

Proof. Assume T to be arcwise connected. Let $f(T)$ be a continuous transformation of T into itself. Let A be any point of T. If $f(A) = B$, where B is not A, then the image of the arc AB, $f(AB)$, is a compact continuous curve intersecting AB in at least B. Indeed, for each positive integer n, $f^n(AB)$ is a compact continuous curve intersecting $f^{n-1}(AB)$. Hence the semi-orbit, H, of AB is a completely separable, connected subset of T, and $\bar{H}$ is a separable subcontinuum of T which is mapped into itself by f. There is a mapping $g(H)$ onto a compact dendrite K satisfying the conditions of Theorem 12, since clearly every increasing sequence of arcs of $\bar{H}$ is contained in an arc of $\bar{H}$. The mapping $gfg^{-1}(K)$ is a continuous mapping of K into itself, for suppose that $X_1, X_2, X_3, \cdots$ is a sequence of points of K converging to a point X. Then it is easy to see that no point of $\bar{H}$ separates $g^{-1}(X_1 + X_2 + X_3 + \cdots)$ from $g^{-1}(X)$ in $\bar{H}$; and hence that no point of $\bar{H}$ separates $fg^{-1}(X_1 + X_2 + \cdots)$ from $fg^{-1}(X)$. It follows that no point of K separates $gfg^{-1}(X_1 + X_2 + \cdots)$ from $gfg^{-1}(X)$ in K, and since this is also true for any subsequence of $X_1, X_2, \cdots$, that $gfg^{-1}(X)$ is a point or a limit point of $gfg^{-1}(X_1 + X_2 + \cdots)$. It follows from the Scherrer Theorem—Theorem 3. 21 of Chapter XII of [25]—that there is a point Y of K such that $gfg^{-1}(Y) = Y$. Then $fg^{-1}(Y) = g^{-1}(Y)$, which proves the theorem.

THEOREM 13(a). *If T is arcwise connected, the condition of Theorem 13 is also necessary.*

Proof. If there is a sequence of arcs of T not satisfying our condition, then it follows that there is a ray R in T. There is a retraction $f(T) = R$ throwing each component of $T - R$ into its boundary point, and there is a "translation" of R, $g(R) = R'$, where R' is a proper subray of R. Then the transformation $gf(T)$ is continuous and leaves no point fixed.

That the condition of Theorem 13 does not imply local compactness or separability, even if T is metric, can be seen by letting T denote the set of all points in or on the unit square, with distance between two points defined as the length of the arc joining the points composed of vertical intervals and, if necessary, an interval of the x-axis. Nor, in general, does it even imply the existence of two non-cutpoints, as will be seen below. Also in general, the condition is not necessary. If α is any ordinal greater than 1, by an α-arc I shall mean the set obtained by taking the sequence of ordinals up to and including α and inserting an open line segment between each two consecutive

ordinals, with the topology defined in a natural way by the order. In this order topology, an α-arc is a bicompact Hausdorff space,[15] and is a tree in the sense of Wallace [18], and hence has the fixed-point property. The set obtained from an α-arc by deleting α, I shall call an α-ray. Some α-rays have no countable sequence running upward through them, and hence have the fixed-point property by Theorem 13, but there are still others that have the property.

THEOREM 14. *If the α-ray R does not contain an increasing sequence $\alpha_1, \alpha_2, \alpha_3, \cdots$ such that no point of R follows all the points of the sequence and such that for $n = 2, 3, \cdots$ the α-arc $\alpha_n\alpha_{n-1}$ is of the same or lower cardinal than $\alpha_{n-1}\alpha_n$, then R has the fixed-point property.*

Proof. Let P be the ordinal 1, and let $T(R)$ be a continuous transformation of R into itself. If P is not fixed, let $T(P) = P'$. There is a point Q of R which is the first point to follow all the points of $T(PP')$, $T^2(PP')$, $T^3(PP'), \cdots$, since otherwise there is a sequence contradicting the condition consisting of one point from each iterate of PP'. The pseudo arc PQ is mapped into itself by T, which, with Wallace's theorem, or Theorem 13, completes the proof.

As another example of the use of Theorem 12, I shall state without proof—which can easily be supplied—a partial extension of a theorem of Schweigert's [16], which has been generalized by Wallace [20].

THEOREM 15. *If T is a generalized dendrite satisfying the condition of Theorem 13, $f(T) = T$ is a homeomorphism, and P is a point of T which is fixed under f and which is an end point of every arc containing it,[16] then there is another point fixed under f.*

4. A fixed-point theorem. In this section we shall combine the results of Sections **2** and **3** to prove that certain connected sets which exhibit dendritic properties have the fixed-point property, even though they are not locally connected. The class of sets we shall consider are the arcwise connected sets such that each increasing sequence of arcs is contained in an arc. Following are several examples of spaces having this property:

a) The sum of the unit interval and a collection of intervals perpendicular to this interval at the points of irrational abscissae.

---

[15] Cf. Theorem 2. 14, p. 28, of [2].

[16] This is not the same as requiring that P be a point of Menger order 1. See the example following Theorem 13.

b) The Cantor star.

c) The continuum obtained by joining the points of a hereditarily indecomposable plane continuum to a point not in the plane by straight-line intervals.

That example b) has the property follows immediately from the result of Hamilton [4] that every hereditarily unicoherent, hereditarily decomposable compact continuum has the fixed-point property, or from the result of Kelley [9] concerning simple links. But neither of these results applies to the other two examples.

THEOREM 16. *Let M be an arcwise connected Hausdorff space which is such that every monotone increasing sequence of arcs is contained in an arc. Then M has the fixed-point property.*

The theorem in the introduction is an immediate consequence.

Proof. In its o-topology, and hence in its a-topology, M contains no simple closed curve; hence in the a-topology it is a generalized dendrite such that every monotone increasing sequence of arcs is contained in an arc. Further, by Theorem 12, every o-continuous mapping of M into a subset, H, of itself is a-continuous in either of the topologies of H determined by the o- or a-topology of M. Hence by Theorem 13, M has the fixed-point property.

The same argument proves from Theorem 15 the following result.

THEOREM 17. *If M is defined as above, and $f(M) = M$ is a homeomorphism, and P is a point fixed under f which is an end point of every arc of M containing it, then there is another point of M fixed under f.*

For the three examples given above, this result does not follow from Wallace's generalization of Schweigert's theorem.

PURDUE UNIVERSITY.

BIBLIOGRAPHY

[1]. Garrett Birkhoff, "On the combination of topologies," *Fundamenta Mathematicae*, vol. 26 (1936), pp. 156-166.

[2]. ———, *Lattice Theory*, American Mathematical Society Colloquium Publications, Vol. XXV, New York, 1940.

[3]. L. M. Blumenthal, *Distance Geometries*, University of Missouri Studies, April, 1938.

[4]. O. H. Hamilton, "Fixed points under transformations of continua which are not

connected im kleinen," *Transactions of the American Mathematical Society*, vol. 44 (1938), pp. 18-24.

[5]. E. Hewitt, "A problem of set-theoretic topology," *Duke Mathematical Journal*, vol. 10 (1943), pp. 309-333.

[6]. F. B. Jones, "Concerning normal and completely normal spaces," *Bulletin of the American Mathematical Society*, vol. 43 (1937), pp. 671-677.

[7]. ———, "Concerning certain linear abstract spaces and simple continuous curves," *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 623-628.

[8]. ———, "Almost cyclic elements and simple links of a continuous curve," *Bulletin of the American Mathematical Society*, vol. 46 (1940), pp. 775-783.

[9]. J. L. Kelley, "A decomposition of compact continua and related theorems on fixed sets under continuous transformations," *Proceedings of the National Academy of Sciences*, vol. 26 (1940), pp. 192-194.

[10]. B. Knaster and C. Kuratowski, "A connected and connected im kleinen point set which contains no arc," *Bulletin of the American Mathematical Society*, vol. 33 (1927), pp. 106-109.

[11]. S. Mazurkiewicz, "Sur les lignes de Jordan," *Fundamenta Mathematicae*, vol. 1 (1920), pp. 106-109.

[12]. K. Menger, "Untersuchungen über allgemeine Metrik. Vierte Untersuchungen. Zur Metrik der Kurven," *Mathematische Annalen*, vol. 103 (1930), pp. 466-501.

[13]. R. L. Moore, "A connected and regular point set which contains no arc," *Bulletin of the American Mathematical Society*, vol. 32 (1926), pp. 331-332.

[14]. ———, *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, Vol. XII, New York, 1932.

[15]. S. B. Myers, "Arcs and geodesics in metric spaces," *Transactions of the American Mathematical Society*, vol. 57 (1945), pp. 217-227.

[16]. G. E. Schweigert, "Fixed elements and periodic types for homeomorphisms on s. l. c. continua," *American Journal of Mathematics*, vol. 66 (1944), pp. 229-244.

[17]. M. E. Shanks, "Monotone decompositions of continua," *American Journal of Mathematics*, vol. 67 (1945), pp. 99-108.

[18]. A. D. Wallace, "A fixed-point theorem for trees," *Bulletin of the American Mathematical Society*, vol. 47 (1941), pp. 757-760.

[19]. ———, "Separation spaces," *Annals of Mathematics*, vol. 42 (1941), pp. 687-697.

[20]. ———, "A fixed-point theorem," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 413-416.

[21]. G. T. Whyburn, "Potentially regular point sets," *Fundamenta Mathematicae*, vol. 16 (1930), pp. 160-172.

[22]. ———, "On the structure of connected and connected im kleinen point sets," *Transactions of the American Mathematical Society*, vol. 32 (1930), pp. 926-943.

[23]. ———, "A junction property of locally connected sets," *American Journal of Mathematics*, vol. 53 (1931), pp. 753-756.

[24]. ———, "A certain transformation on metric spaces," *American Journal of Mathematics*, vol. 54 (1932), pp. 367-376.

[25]. ———, *Analytic Topology*, American Mathematical Society Colloquium Publications, Vol. XXVIII, New York, 1942.

[26]. G. S. Young, "On continua whose links are non-intersecting," *Bulletin of the American Mathematical Society*, vol. 50 (1944), pp. 920-925.

ON THE SOLUTIONS OF AN ORDINARY DIFFERENTIAL EQUATION NEAR A SINGULAR POINT.*

By Philip Hartman.

This paper is a continuation of the considerations of a recent paper by Wintner and the present author [3] dealing with Perron's [4] generalization of certain of Poincaré's qualitative results on the singularity $(x, y) = (0, 0)$ of the real (analytic) differential equation

$$(1) \qquad xy' = \alpha x + \beta y + \cdots \qquad (\beta \neq 0)$$

and the corresponding results of Bendixson on the non-analytic equation

$$x^m y' = \alpha x + \beta y + \cdots \qquad (\beta \neq 0,\ m = 1, 2, \cdots)$$

(cf. Liebmann, [2], pp. 507-512 and Dulac [1], p. 178 for further references).

Perron dealt with the differential equation

$$(2) \qquad \phi(x) y' = f(x, y),$$

where $\phi(x)$ is any positive continuous function on an interval $0 < x \leqq a$ satisfying

$$(3) \qquad \phi(x) \rightarrow 0 \text{ as } x \rightarrow +0$$

and

$$(4) \qquad \int_{+0} dx/\phi(x) = +\infty,$$

and where $f(x, y)$ is a real-valued continuous function on a closed rectangle $0 \leqq x \leqq a$, $|y| \leqq b$ subjected to the condition that

$$(5) \qquad f(0, 0) = 0$$

and to the upper and lower Lipschitz conditions

$$(6) \qquad |(f(x, y_1) - f(x, y_2))/(y_1 - y_2)| < C, \qquad (y_1 \neq y_2),$$

$$(6\text{ bis}) \qquad |(f(x, y_1) - f(x, y_2))/(y_1 - y_2)| > c > 0, \qquad (y_1 \neq y_2).$$

Perron showed that there are two different situations depending on the algebraic sign of the non-vanishing difference quotient in (6 bis). If this

* Received April 17, 1946.

sign is plus, there exist two positive numbers a', b' such that every solution path $y=y(x)$ of (2) issuing from a point of the rectangle

$$R' : 0 < x < a', |y| < b' \qquad (a' \leqq a, b' \leqq b) \tag{7}$$

satisfies

$$\lim_{x\to+0} y(x) = 0. \tag{8}$$

If the sign is minus, there exists exactly one solution path $y=y(x)$ satisfying (8).

In [3], it was shown that Perron's analytic proof involving successive approximations could be avoided by the use of simple geometrical considerations, and at the same time, the heavy conditions imposed by Perron could be lightened to a great degree. The principal theorem proved there was:

(*) *Let* $\phi(x)$, *where* $0 < x \leqq a$, *be a positive, continuous function satisfying* (4). *Let* $f(x,y)$ *be defined and continuous on the (partly open) rectangle*

$$R: 0 < x \leqq a, |y| \leqq b, \tag{9}$$

having the property that $1/f(x,y)$, *outside of any fixed vicinity of the origin, is bounded for small* x *and that* $f(x,b)$ *and* $f(x,-b)$ *do not vanish for* $0 < x \leqq a$ *and are of opposite signs.*

(I) *if, further,* $f(x,b) > 0$ *and* $f(x,-b) < 0$ *for* $0 < x \leqq a$, *then every solution path* $y=y(x)$ *of* (2) *issuing from a point* (x_0,y_0) *of the rectangle* (9) *can be continued for all positive* $x(< x_0)$ *and* (8) *holds for all continuations.*

(II) *If, however,* $f(x,b) < 0$ *and* $f(x,-b) > 0$ *for* $0 < x \leqq a$, *then there exists at least one solution path* $y=y(x)$ *of* (2) *defined for all small positive* x *and* (8) *holds for all such paths.*

(It may be remarked, for completeness' sake, that the solution in (II) is unique, if $f(x,y)$ is, for every fixed x, monotone with respect to y).

In addition to the limit relation (8), Perron also considered the behavior of the ratio $y(x)/x$. Under the conditions imposed by him on $f(x,y)$, there exists, for small $x \geqq 0$, a unique continuous function $y=y^0(x)$ such that $y^0(0)=0$ and

$$f(x,y^0(x)) \equiv 0. \tag{10}$$

Assuming further that

$$\lim_{x\to+0} y^0(x)/x = \gamma \text{ exists;} \tag{11}$$

and that

$$\phi(x)/x \to 0, \text{ as } x \to +0, \tag{12}$$

Perron showed that for any solution path $y=y(x)$ of (2) satisfying (8),

$$\lim_{x\to+0} y(x)/x=\gamma. \tag{13}$$

The object of this paper is to obtain similar relations under improved conditions by the use of the geometrical arguments employed in [3]. (For another investigation of similar asymptotic statements in the non-analytic case of (1), cf. Wintner [6]; the results of this paper are in a different direction, however.)

In the case of the above italicized theorem (*), it turned out that Perron's (upper) Lipschitz condition (6) was entirely superfluous, while the (lower) condition (6 bis) could be considerably weakened to conditions on $f(x,y)$ near the boundary lines $x=0$, $y=+b$, $y=-b$. In the present case, the situation is similar in that (6) is not needed, while (6 bis) is retained either in its full strength or, in some instances, in a weakened form.

First, the following theorem will be proved:

THEOREM 1. *Let $\phi(x)$ be a positive, continuous function on the interval $0<x\leqq a$ satisfying* (12). *Let $f(x,y)$ be a real-valued, continuous function on the partly open rectangle R,* (9), *and satisfy there the lower Lipschitz condition* (6 bis). *Let there exist a function $y=y^0(x)$, $0<x\leqq a$, satisfying $|y^0(x)|<b$, and* (10) *and* (11). *If $y=y(x)$ is any solution path of* (2) *satisfying* (8), *then* (13) *holds.*

Proof. It may be remarked that as far as the existence of solutions $y=y(x)$ of (2) satisfying (8) is concerned, the conditions of the theorem (*) above are satisfied. In fact, for the applicability of (*), $y^0(x)\to 0$ as $x\to+0$ could replace (11).

It may be supposed that the constant c in (6 bis) is 1; otherwise the differential equation (2) can be divided by c and this factor absorbed in both $f(x,y)$ and $\phi(x)$ without changing either the conditions or the statement of the theorem. Also, in order to fix ideas, it will be supposed that the algebraic sign of the non-vanishing difference quotient in (6 bis) is plus. The proof in the other case is similar. Thus, (6 bis) becomes

$$f(x,y_1)-f(x,y_2)>y_1-y_2 \qquad \text{if } y_1>y_2, \tag{14}$$

for $0<x\leqq a$.

Introducing the new dependent variable

$$v=v(x)=y(x)/x, \tag{15}$$

the differential equation (2) becomes

(16) $$\phi(x)v' = F(x, v),$$
where
(17) $$F(x, v) = [f(x, vx) - v\phi(x)]/x$$

is continuous on its domain of definition $0 < x \leqq a$, $|v| \leqq b/x$. This function satisfies the inequality

(18) $$F(x, v_1) - F(x, v_2) > (v_1 - v_2)(1 - \phi(x)/x) \text{ if } v_1 > v_2.$$

It follows from (12) that for sufficiently small positive x, $F(x, v)$ is monotone increasing in v.

It will be shown *that if x is sufficiently small, there exists a unique value of $v = v^0(x)$ such that $F(x, v^0(x)) = 0$ and that $v^0(x)$ lies between $y^0(x)/x$ and $y^0(x)/(x - \phi(x))$*. That $v^0(x)$ is unique follows from the strict monotony of $F(x, v)$ with respect to v. Now, place $v = y^0(x)/x$ in the definition (17) of $F(x, v)$; by (10)

$$F(x, y^0(x)/x) = -y^0(x)\phi(x)/x^2$$

which either vanishes (together with $y^0(x)$) or has the same sign as $-y^0(x)$. If $y^0(x) = 0$, nothing remains to be proved. Suppose, for a moment, that $y^0(x) > 0$ for this fixed x; so that $xy^0(x)/(x - \phi(x)) > y^0(x)$ for small $x > 0$. Subtracting $0 = f(x, y^0(x))/x$ from the function (17), with v replaced by $y^0(x)/(x - \phi(x))$, and applying (14) to $y_1 = xy^0(x)/(x - \phi(x))$ and $y_2 = y^0(x)$, one obtains

$$F(x, y^0(x)/(x - \phi(x)) > 0.$$

Similarly, if it is supposed that $y^0(x) < 0$, it follows that $F(x, y^0(x)/(x - \phi(x))) < 0$. In view of the continuity of $F(x, v)$, this concludes the proof of the italicized statement concerning $v^0(x)$.

It follows from (11) and (12) that

(19) $$v^0(x) \to \gamma \text{ as } x \to +0.$$

From (18) and (19) it also follows that if a' is any sufficiently small positive number and if $b' = b/a'$, then $F(x, b') > 0$ and $F(x, -b') < 0$ for $0 < x \leqq a'$ and that $1/F(x, v)$, outside of any neighborhood of the point $(x, v) = (0, \gamma)$, is bounded for small $x > 0$ (when $|v| \leqq b'$). The proof of (*) implies that if $v = v(x)$ is any solution path of (16) through a point (x_0, v_0) of the rectangle $0 < x_0 \leqq a'$, $|v_0| \leqq b'$, then $v(x)$ can be continued for all positive $x (< x_0)$ and for all continuations

(20) $$v(x) \to \gamma \text{ as } x \to +0.$$

In view of the fact that every solution path $y = y(x)$ of (2) satisfying (8) corresponds to a solution $v = v(x)$ of (16) and that $|v(x_0)| = |y(x_0)/x_0| \leqq b/x_0$ for all $x = x_0$, it can be concluded, by choosing $a' = x_0$ (and $b' = b/a'$) for sufficiently small x_0, that (20) holds.

By (15), this proves Theorem 1.

From the proof of this theorem and from the proof of (*), the following corollary may be deduced:

COROLLARY 1. *Let $\phi(x)$ be a positive, continuous function in the interval $0 < x \leqq a$ such that*

(21) $$\limsup_{x \to +0} \phi(x)/x = \lambda < +\infty.$$

Let $f(x, y)$ satisfy the conditions of Theorem 1. Let

(22) $$\lambda < c,$$

where $c > 0$ is a constant satisfying (6 bis). If $y = y(x)$ is any solution path of (2) satisfying (8), then $\liminf y(x)/x$ *and* $\limsup y(x)/x$ *lie between γ and $c\gamma/(c - \lambda)$.*

The proof of this corollary merely depends on the fact that, in the proof of Theorem 1, where it was supposed that $c = 1$, one can conclude only that $\limsup v^0(x)$ and $\liminf v^0(x)$ lie between γ and $\gamma/(1 - \lambda)$ if (21) replaces (12).

If the limit γ in (11) is zero, this corollary becomes:

COROLLARY 2. *Let $\phi(x)$ be a positive, continuous function on the interval $0 < x \leqq a$ satisfying (21). Let $f(x, y)$ satisfy the conditions of Theorem 1 with (11) particularized to*

(11 bis) $$\lim_{x \to +0} y^0(x)/x = 0.$$

Let (22) be satisfied. If $y = y(x)$ is any solution path of (2) satisfying (8), then

(13 bis) $$\lim_{x \to +0} y(x)/x = 0.$$

It may be remarked that condition (6 bis) in Theorem 1 cannot be replaced by the condition that $f(x, y)$ be strictly monotone with respect to y for a fixed $x > 0$ (a condition which suffices for (*)). This may be seen from the example

$$x^2 y' = y^3$$

where the solutions are $y \equiv 0$ or $y = [x/2(1 + Kx)]^{\frac{1}{2}}$, K being an arbitrary

integration constant. Thus (13 bis) fails to hold except for the solution $y \equiv 0$ even though (11 bis) and (12) are true.

Nevertheless, Corollary 2 can be improved in some directions; the first extension [1] corresponding to the case (II) of (*) is:

THEOREM 2. *Let $\phi(x)$ and $f(x,y)$ satisfy the conditions of (*) in the case* (II). *Let $y^+(x)$, $y^-(x)$ denote, for a fixed $x > 0$, the greatest and least value of y, $|y| < b$, for which $f(x,y) = 0$, and let*

$$\lim_{x\to 0} y^+(x)/x = \lim_{x\to 0} y^-(x)/x = 0. \tag{23}$$

Then (13 bis) *holds for any solution path $y = y(x)$ of* (2) *satisfying* (8).

Proof. Let $\epsilon > 0$ be fixed and let $\delta = \delta(\epsilon)$ be so small that

$$|y^+(x)| < \epsilon x \text{ and } |y^-(x)| < \epsilon x \text{ if } 0 < x < \delta.$$

Then, if $y = y(x)$ is any solution path of (2) satisfying (8),

$$|y(x)| < \epsilon x \text{ if } 0 < x < \delta. \tag{24}$$

For suppose, if possible, that for some $x = \bar{x}$, $0 < \bar{x} < \delta$,

$$y(\bar{x}) \geqq \epsilon\bar{x} > 0.$$

In particular $y(\bar{x}) > y^+(\bar{x})$, and so from the definition of y^+, the continuity of $f(x,y)$ and the fact that $f(x,b) < 0$, it follows that $f(\bar{x}, y(\bar{x})) < 0$. From (2), it is seen that $y'(\bar{x}) < 0$ and so $y(x)$ increases with decreasing x at $x = \bar{x}$. This situation must pertain for all x, $0 < x \leqq \bar{x}$. Since $y(\bar{x}) > 0$, the relation (8) cannot hold. This contradiction proves (24).

This completes the proof of Theorem 2.

For the corresponding case (I) of (*), the above example indicates that more stringent conditions are needed. There will now be proved: [2]

THEOREM 3. *Let $\phi(x)$ and $f(x,y)$ satisfy the conditions of (*) in the case* (I). *Let, in addition,*

---

[1] In [5], Perron proved a similar theorem, namely: *If $f(x,y)$ $g(x,y)$ are continuous on the closed rectangle $0 \leqq x \leqq a$, $|y| \leqq b$ and $f(x,y) = o(|x| + |y|)$, $g(x,y)$ $o(|x| + |y|)$ as $x, y \to 0$, then there exists at least one solution path $y = y(x)$ of*

$$[x + g(x,y)]dy = -c[y + f(x,y)]dx, \qquad c > 0,$$

such that $y(x) \to 0$ and $y'(x) \to 0$ as $x \to +0$.

[2] This theorem is an improvement over that of Perron [5], Satz 3, p. 129: *If $g(x,y)$ is continuous on the closed rectangle $0 \leqq x \leqq a$, $|y| \leqq b$ and $g(x,y) = o(x + |y|)$ as $x, y \to 0$, then for any solution $y = y(x)$ of*

$$xy' = cy + g(x,y), \qquad c > 1,$$

satisfying (8), *one has* (13 bis).

$$(25) \qquad |f(x,y)| > c\,|y| + o(x) \text{ as } x, y \to 0 \qquad (c > 0).$$

Finally, let $\phi(x)$ *satisfy* (21) *and let* (22) *be satisfied. Then, for any solution path* $y = y(x)$ *of* (2), (13 bis) *holds.*

Proof of Theorem 3. If $y^+(x)$ and $y^-(x)$ have the same meaning as in Theorem 2, (25) implies (23).

Now, introduce the dependent variable (15), so that (2) becomes the equation (16). Let y be a fixed number such that $y > \max(0, y^+(x))$. The function $xF(x, y/x) = f(x,y) - y\phi(x)/x$ is not less than

$$(26) \qquad y(c - \phi(x)/x) + o(x) \qquad \text{as } x, y \to 0$$

by (25). From (21) and (22), it is seen that (26) is a positive quantity for sufficiently small $x > 0$. Similarly, if $y < \min(0, y^-(x))$, then $xF(x, y/x)$ is not greater than (26), which is negative for sufficiently small $x > 0$. Consequently, there exist two positive numbers a', b' ($a' \leqq a$, $b' \leqq b$) such that for every x, $0 < x \leqq a'$, there exists at least one value of y, $|y| \leqq b'$, for which $xF(x, y/x) = 0$; and if $y_1^+(x)$, $y_1^-(x)$ denotes the greatest and least such values of y, then

$$y_1^+(x)/x \to 0, \qquad y_1^-(x)/x \to 0, \text{ as } x \to +0.$$

Hence, for every x, $0 < x \leqq a'$, there exists at least one value of v, $|v| \leqq b'/x$, for which $F(x, v) = 0$; and if $v^+(x)$, $v^-(x)$ denote the greatest and least such values of v, then

$$v^+(x) \to 0, \; v^-(x) \to 0 \text{ as } x \to +0$$

(actually $v^- = y_1^+/x$; $v^- = y_1^-/x$).

Also, from (17) and (25),

$$|F(x, v)| > (c - \phi(x)/x)\,|v| + o(x) \text{ as } x, |vx| \to 0.$$

The proof of Theorem 3 can now be completed with the same arguments used in the proof of Theorem 1.

These two theorems yield the following corollaries:

Corollary 3. *Let* $f(x, y)$ *be a continuous function on the partly open rectangle* R, (9), *and let*

$$(27) \qquad f(x,y) = \alpha x + \beta y + o(x + |y|) \text{ as } x, y \to 0$$

where

$$(28) \qquad \beta < 0.$$

Let $\phi(x)$ *be a positive, continuous function on the interval* $0 < x \leqq a$ *and let*

$$\phi(x)/x \to \lambda \text{ as } x \to +0. \tag{29}$$

If $y = y(x)$ *is any solution of* (2) *satisfying* (8), *then*

$$\lim_{x\to+0} y(x)/x = -\alpha/(\beta - \lambda). \tag{30}$$

COROLLARY 4. *Let* $f(x, y)$ *be a continuous function on the partly open rectangle* R, (9), *and let* (27) *hold, but*

$$\beta > 0.$$

Let $\phi(x)$ *satisfy the conditions of Corollary* 3 *and*

$$\lambda < \beta. \tag{31}$$

Then there exist two positive numbers a', b' *such that if* $y = y(x)$ *is a solution of* (2) *issuing from a point* (x_0, y_0) *of the rectangle* R', (7), *then* $y = y(x)$ *can be continued for all positive* $x (< x_0)$ *and* (30) *holds for any such continuation.*

Proof of Corollaries 3 *and* 4. Introducing the new dependent variable

$$w = w(x) = y(x) + \alpha x/(\beta - \lambda),$$

the differential equation (2) becomes

$$\phi(x) w' = g(x, w), \tag{32}$$

where the function

$$g(x, w) = f(x, w - \alpha x/(\beta - \lambda)) + \alpha\phi(x)/(\beta - \lambda)$$

is continuous on the partly open parallelogram $0 < x \leqq a$, $| w - \alpha x/(\beta - \lambda) | \leqq b$. From (27) and (29),

$$g(x, w) = \beta w + o(x + | w - \alpha x/(\beta - \lambda) |)$$

as $x \to 0$ and $w - \alpha x/(\beta - \lambda) \to 0$, which implies

$$g(x, w) = \beta w + o(x + | w |) \text{ as } x,\ w \to 0.$$

Corollaries 3 and 4 now follow by an application of Theorems 2 and 3, respectively, to the equation (32).

If conditions of the type (21) and (22) or (29) and (31) on $\phi(x)$ are weakened to allow the "$<$" to be replaced by "$\leqq$," it is easy to see that in general no statement concerning the existence of the limit of $y(x)/x$ can be made. However, even without assumptions on the zeros of $f(x, y)$, one can

obtain results about the ratio $(y_1(x)-y_2(x))/x$, where $y=y_1(x)$ and $y=y_2(x)$ are two solutions of (2). The first of such statements is:

THEOREM 4. *Let $\phi(x)$ and $f(x,y)$ satisfy the conditions of (*) in the case* (I). *Let, in addition, $f(x,y)$ satisfy the lower Lipschitz condition* (6 bis). *Let $\phi(x)$ satisfy*

$$(33)\qquad c/\phi(x)-1/x \geqq 0 \textit{ for small } x>0,$$

where c is a constant satisfying (6 bis). *Then, if $y=y_1(x)$ and $y=y_2(x)$ are any two solutions of* (2) *satisfying* (8),

$$(34)\qquad \lim_{x\to+0}(y_1(x)-y_2(x))/x \textit{ exists}$$

($\pm\infty$ *are allowed in* (34)).

Proof. Again assume, for simplicity, that $c=1$, so that (6 bis) becomes (14). Introduce the dependent variable (15), so that the equation (2) becomes (16).

Let $y=y_1(x)$, $y=y_2(x)$ be two solutions of (2) satisfying (8). It may be supposed that $y_1(x)\geqq y_2(x)$ for all small x (otherwise the indices may be interchanged for some x). If $v_1=y_1(x)/x$, $v_2=y_2(x)/x$, it follows from (16) and (18) that

$$(36)\qquad (v_1-v_2)'\geqq(v_1-v_2)[1/\phi(x)-1/x].$$

From (33) it is seen that $(v_1-v_2)'\geqq 0$, so that v_1-v_2 is a monotone function for small x. (34) now follows.

In analogy with (*), one might expect that the conditions (21) and (22) or (29) and (31) can be replaced by an integral condition similar to (4); for example, a condition like

$$(37)\qquad \int_{+0}(c/\phi(x)-1/x)\,dx=+\infty.$$

However, the solutions of the differential equation

$$(1/x-1/x\log x)^{-1}y'=y+x$$

are

$$y=\tfrac{1}{2}x\log x-x+Kx/\log x,$$

where K is an arbitrary integration constant; so that it is false that (37) is sufficient for the existence of a limit of the ratio $y(x)/x$. But (37) may be used to improve Theorem 4 as follows:

THEOREM 5. *Let $\phi(x)$, $f(x,y)$ satisfy the conditions of Theorem* 4.

In addition, let (37) *be satisfied. If* $y = y_1(x)$, $y = y_2(x)$ *are any two solutions of* (2) *satisfying* (8), *then*

$$\lim_{x \to +0} (y_1(x) - y_2(x))/x = 0. \tag{38}$$

Proof. Let y_1, y_2, v_1, v_2 have the same meaning as in the proof of Theorem 4, so that (36) holds. Consider the differential equation

$$u' = u[1/\phi(x) - 1/x].$$

It is clear from (37) that for any solution

$$u = u(x) = (\text{const.}) \exp\left(- \int_x [1/\phi(t) - 1/t]dt\right),$$

$$u(x) \to 0, \; x \to +0. \tag{39}$$

Now if $v_1(x) - v_2(x)$ and $u(x)$ have a common value for some x, then, by (36), for smaller (positive) x,

$$v_1(x) - v_2(x) \leqq u(x).$$

The relation (39) and the fact that $v_1(x) - v_2(x) \geqq 0$ imply (38).

This completes the proof of Theorem 5.

QUEENS COLLEGE, FLUSHING, N. Y.

REFERENCES

[1] H. Dulac, *Curvas definidas por una ecuación deferencial de primer orden y de primer grado*, Madrid, 1933.

[2] H. Liebmann, "Geometrische Theorie der Differentialgleichungen," *Encyklopaedie der mathematischen Wissenschaften*, III D 8 (1914).

[3] P. Hartman and A. Wintner, "On the asymptotic behavior of the solutions of a non-linear differential equation," *American Journal of Mathematics*, vol. 68 (1946), pp. 301-308.

[4] O. Perron, "Beweis für die Existenz von Integralen einer gewöhnlichen Differentialgleichung in der Umgebung einer Unstetigkeitsstelle." *Mathematische Annalen*, vol. 75 (1914), pp. 256-273.

[5] O. Perron, "Über die Gestalt der Integralkurven einer Differentialgleichung erster Ordnung in der Umgebung einer singulären Punktes," *Mathematische Zeitschrift*, vol. 15 (1922), pp. 121-146 and vol. 16 (1923), pp. 273-295.

[6] A. Wintner, "Asymptotic integration constants in the singularity of Briot-Bouquet," *American Journal of Mathematics*, vol. 68 (1946), pp. 293-300.

LINEAR GRAPHS OF DEGREE ≤ 6 AND THEIR GROUPS.*

By I. N. KAGNO.

1. Introduction. Let G be a connected linear graph having no simple loops (i. e. no arcs with coincident endpoints, or pairs of arcs which form a circuit) and each vertex of which is of degree ≥ 3. Any one-one, continuous map of G into itself effects a permutation among some or all of its vertices. If we denote the vertices of G by $a_1, \cdots, a_n$ then, corresponding to each map T of G into itself, there is a substitution τ on the letters $a_1, \cdots, a_n$. We shall say that G is mapped into itself by the substitution τ. Thus, if Γ denotes the group of all possible maps of G into itself, the set of corresponding substitutions constitute a substitution group $\mathfrak{G}$. We shall say that a given graph G *has the group* $\mathfrak{G}$, or that $\mathfrak{G}$ *is the group of* G, if corresponding to every map of G into itself there is a substitution of $\mathfrak{G}$, and each substitution of $\mathfrak{G}$ represents a map of G into itself. We shall also say that $\mathfrak{G}$ *has the graph* G, or that G *is the graph of* $\mathfrak{G}$.

In this paper we seek to determine whether a given graph of Degree ≤ 6 has a non-identical group (where by the Degree of a graph we mean the number of its vertices; to avoid confusion between the words *Degree* as applied to graphs and *degree* as applied to vertices, we shall use a capital D in the former case and a lower case d in the latter) and whether a given group of degree ≤ 6 has a graph.[1] We shall show that every graph of Degree ≤ 6 has a non-identical group. The converse of this is not true, for as we shall show, there exist groups of any degree > 2 which do not have graphs. We shall also give an example of a graph that does not have any non-identical group. Thus in general, not every graph has a non-identical group, nor does every group have a graph. We shall find the group for each graph of Degree ≤ 6, and examine each group of degree ≤ 6 to determine whether it has a graph.

* Received January 4, 1946; Revised May 6, 1946.

[1] The second question is a special case of the more general question proposed by König in his book "Theorie der endlichen und unendlichen Graphen," Leipzig (1936): "When can a given abstract group be set up as the group of a graph, and if possible how can the graph be constructed?" The graphs referred to by König may also contain simple loops.

I.

2. Some theorems on graphs of any Degree.

Definition. Let $I_b{}^a$ be a number defined for a graph G as follows; if G contains the arc $\widehat{ab}$ then $I_b{}^a = I_a{}^b = 1$. If G does not contain the arc $\widehat{ab}$ then $I_b{}^a = I_a{}^b = 0$. $I_b{}^a$ is called the *adjacency number* of the pair of vertices a, b. If G has the property that G contains the arc $\widehat{rs}$ if and only if it contains the arc $\widehat{pq}$, then we have $I_q{}^p = I_s{}^r$. If G is mapped into itself by a substitution τ, then an adjacent pair of vertices must be carried into a pair that is adjacent, and a non adjacent pair of vertices cannot be carried into an adjacent pair. In other words, τ is arc-preserving. Hence if a, b, p, q are vertices of G such that $\tau(a) = p$, $\tau(b) = q$, then $I_b{}^a = I_q{}^p$. Similarly if τ leaves a fixed and $\tau(b) = q$ then $I_b{}^a = I_q{}^a$.

THEOREM 1. *Let τ be a substitution on the vertices of a graph G such that $I_q{}^p = I_{\tau(q)}{}^{\tau(p)}$ for every pair of vertices p, q, of G. Then there exists a one-one continuous mapping σ of G into itself such that $\sigma(x) = \tau(x)$ for every vertex x of G. We shall say that τ maps G into itself.*

Proof.[2] Let $\widehat{pq}$ be any arc of G. Then, by hypothesis, G contains $\widehat{p'q'}$, where $p' = \tau(p)$, $q' = \tau(q)$. If p'', q'' are vertices such that $p'' = \tau^{-1}(p)$, $q'' = \tau^{-1}(q)$ then, by hypothesis, $I_{q''}{}^{p''} = I_{\tau(q'')}^{\tau(p'')} = I_q{}^p$; hence G also contains $\widehat{p''p''}$.

Let σ be a single-valued transformation of G into itself defined as follows: Suppose that P is any point of G. (a) If P is a vertex then we set $\sigma(P) = \tau(P)$. (b) If P is an interior point of an arc $\widehat{pq}$, let T_{pq} be a continuous one-one transformation of the closed arc $\widehat{pq}$ onto the closed arc $\widehat{p'q'}$, such that $T_{pq}(p) = p'$, $T_{pq}(q) = q'$. Let $\sigma(P) = T_{pq}(P)$.

Since T_{pq} is defined and continuous on the closure of $\widehat{pq}$, and $\sigma(p) = \tau(p) = p' = T_{pq}(p)$, $\sigma(q) = \tau(q) = q' = T_{pq}(q)$, it is clear that σ is a continuous transformation which carries every point of G into a unique point of G.

Let σ^* be a single-valued transformation of G into itself defined as follows; if P is a vertex of G let $\sigma^*(P) = \tau^{-1}(P)$, and if P is an interior point of an arc $\widehat{pq}$, let $\sigma^*(P) = T_{pq}{}^{-1}(P)$. Then as above, σ^* is continuous and carries every point of G into a unique point of G. It is readily seen that σ^*

[2] The proof which follows, an improvement over the author's original proof, was suggested by the referee. The author wishes to thank the referee for this and other valuable suggestions in the preparation of this paper for publication.

and σ are inverses of each other. Hence σ is one-one, as well as continuous. By definition $\sigma(x) = \tau(x)$ for every vertex x. Obviously σ preserves incidence. Hence σ satisfies the conditions of the theorem.

THEOREM 2. *If a graph G of Degree α_0 has a group $\mathfrak{G}$, a necessary condition that $\mathfrak{G}$ be of degree α_0 is that if a_i is a vertex of degree δ of G, then G contains at least one other vertex of degree δ.*

Proof. Suppose that a_i is of degree δ, and that G does not contain any other vertex of degree δ. If every map T of G into itself leaves a_i fixed, then the group of G cannot be of degree α_0, contrary to hypothesis. Hence let T be a map which does not leave a_i fixed and set $T(a_i) = a_j$. Then the arcs incident with a_i are mapped into arcs incident with a_j, any two distinct arcs on the former going into two distinct arcs on the latter. If the degree of a_j is less than δ this is not possible. If the degree of a_j exceeds δ, then there are arcs on a_j which are not the image of any arcs on a_i. Since T is one-one this is impossible. Hence G must contain another vertex of degree δ.

We shall call the group of substitutions on n letters, which transforms a plane n-gon into itself the *n-gonal group.*

THEOREM 3. *The complete N-point has the group $\mathfrak{S}_n \equiv (a_1 a_2 \cdots a_n)$* all.
As this theorem is well known,[3] we omit the proof.

THEOREM 4. *If a graph G is mapped into itself by the substitution $\sigma = (a_1 a_2 \cdots a_n)$ on n of its vertices $a_1, a_2, \cdots, a_n$, then G is also mapped into itself by each substitution of the n-gonal group $\mathfrak{N}$ on the letters $a_1, \cdots, a_n$.*

Proof. $\sigma = (a_1 a_2 \cdots a_n)$ leaves fixed every vertex of G distinct from the vertices a_j, $(j = 1, \cdots, n)$. Let $p \neq a_j$ be a vertex of G adjacent to a_1. Since σ leaves p fixed, $I_{a_1}{}^p = \cdots = I_{a_n}{}^p = 1$. Similarly if q is any vertex of G distinct from $a_1, \cdots, a_n$,

$$I_{a_1}{}^q = \cdots = I_{a_n}{}^q. \tag{1}$$

Suppose that $I^{a_2}_{a_1+i} = 1$, $(i = 1, \cdots, n-1)$. Then $I^{\sigma(a_1)}_{\sigma(a_1+i)} = 1$. Similarly if $I^{a_1}_{a_1+i} = 0$, then $I^{\sigma(a_1)}_{\sigma(a_1+i)} = 0$. That is $I^{a_1}_{a_1+i} = I^{\sigma(a_1)}_{\sigma(a_1+i)} = I^{a_2}_{a_2+i}$ (the additions in the subscripts being understood to be modulo n). Analogously

$$I^{a_1}_{a_1+i} = I^{a_2}_{a_2+i} = \cdots = I^{a_n}_{a_n+i}. \tag{2}$$

[3] Cf. R. Frucht, "Die Gruppe des Peterschen Graphen etc.," *Com. Math. Helv.*, vol. 9 (1937), pp. 217-223.

Since σ maps G into itself, G is also mapped into itself by each of the substitutions $\sigma^2, \sigma^3, \cdots, \sigma^{n-1}$.

To show that the remaining substitutions of $\mathfrak{R}$ map G into itself, we must consider two cases; 1.) n is odd, 2.) n is even.

Case 1. n is odd. Consider the substitution

$$\tau = (a_2a_n)(a_3a_{n-1}) \cdots (a_ja_{n+2-j}) \cdots (a_{(n+1)/2}a_{(n+3)/2}).$$

τ leaves fixed any pair of vertices p, q distinct from a_j, $(j = 1, \cdots, n)$, and hence any arc $\widehat{pq}$. That is, $I_q{}^p = I_{\tau(p)}^{\tau(q)}$. From (1), if p is any vertex distinct from a_j, then $I_{a_j}{}^p = I_{a_{n+2-j}}^{p} = I_{\tau(a_j)}^{\tau(p)}$. If a_j and a_k are distinct, let $k = j + i$. Then from (2),

$$I_{a_k}^{a_j} = I_{a_{j+i}}^{a_j} = I_{a_{(n+2-k)+i}}^{a_{(n+2-k)}} = I_{a_{n+2-j}}^{a_{n+2-k}} = I_{a_{n+2-k}}^{a_{n+2-j}} = I_{\tau(a_k)}^{\tau(a_j)}.$$

Thus for any pair of vertices u, v of G, $I_v{}^u = I_{\tau(v)}^{\tau(u)}$. Hence by Theorem 1 τ maps G into itself.

Now any substitution of $\mathfrak{R}$ is a product of a power of τ and a power of σ (for σ and τ are the generators of the $2n$ substitutions of $\mathfrak{R}$). Hence any such substitution maps G into itself.

Case 2. n is even. Consider the substitution

$$\tau = (a_1a_n)(a_2a_{n-1}) \cdots (a_ja_{n+1-j}) \cdots (a_{n/2}a_{(n+2)/2}).$$

τ leaves fixed any pair of vertices p, q distinct from a_j. That is $I_q{}^p = I_{\tau(q)}^{\tau(p)}$. From (1), if p is any vertex distinct from a_j, $I_{a_j}{}^p = I_{\tau(a_j)}^{\tau(p)}$. If a_j and a_k are distinct, then from (2)

$$I_{a_k}^{a_j} = I_{a_{(n+1-k)+i}}^{a_{(n+1-k)}} = I_{a_{n+1-j}}^{a_{n+1-k}} = I_{a_{n+1-k}}^{a_{n+1-j}} = I_{\tau(a_k)}^{\tau(a_j)}.$$

Thus for any pair of vertices u, v of G, $I_v{}^u = I_{\tau(v)}^{\tau(u)}$. Hence by Theorem 1, τ maps G into itself.

In an analogous manner we can show that the substitution

$$\rho = (a_2a_n)(a_3a_{n-1}) \cdots (a_ja_{n+2-j}) \cdots (a_{n/2}a_{(n+4)/2})$$

maps G into itself.

Since $\mathfrak{R}$ is generated by the substitutions σ, τ, and ρ,[4] any substitution of $\mathfrak{R}$ maps G into itself.

---

[4] Cf. Miller, Blichfeldt and Dickson, *Theory and Applications of Finite Groups*, p. 9.

COROLLARY 1. *The cyclic group* $(abc\cdots n)\mathrm{cyc}$ *on* n *letters,* $(n>2)$ *does not have a graph.*

Proof. Suppose that G is the graph of the group $(abc\cdots n)\mathrm{cyc}$. Then G is mapped into itself by the substitution $\sigma=(abc\cdots n)$. But then by the Theorem, G is mapped into itself by the substitution $\tau_1=(a_2a_n)\cdots(a_ja_{n+2-j})\cdots(a_{(n+1)/2}a_{(n+3)/2})$ when n is odd, or $\tau_2=(a_1a_n)\cdots(a_ja_{n-j+1})\cdots(a_{n/2}a_{(n+2)/2})$ when n is even. Since neither of these substitutions is contained in the given group this gives a contradiction.

COROLLARY 2. *The group* $(a_1a_2\cdots a_n)\mathrm{cyc}(pq\cdots s)$ *does not have any graph.*

Proof. Suppose that the given group has the graph G. Then G is mapped into itself by the substitution $(a_1a_2\cdots a_n)$ on n of its vertices $a_1,a_2,\cdots,a_n$, for the given group is the direct product of the groups $\mathfrak{G}_1\equiv(a_1a_2\cdots a_n)\mathrm{cyc}$ and $\mathfrak{G}_2\equiv[1,(pqr\cdots s)]$. But by the Theorem, if n is odd G is then mapped into itself by the substitution $\tau_1=(a_2a_n)\cdots(a_ja_{n+2-j})\cdots(a_{(n+1)/2}a_{(n+3)/2})$ and if n is even G is mapped into itself by the substitution $\tau_2=(a_1a_n)\cdots(a_ja_{n-j+1})\cdots(a_{n/2}a_{(n+2)/2})$. Since neither τ_1 nor τ_2 are contained in the given group, this gives a contradiction.

THEOREM 5. *The substitution group* $(abc\cdots m)\mathrm{pos}$ *does not have any graph.*

Proof. Since the group contains only positive substitutions, it does not contain any transpositions. Furthermore since the group contains all positive substitutions on m letters, it contains in particular the substitution (abc). Now suppose that the group has the graph G. Then G is mapped into itself by the substitution (abc) on three of its vertices a,b,c. But then G is mapped into itself by all substitutions of the 3-gonal group $(abc)\mathrm{all}$, (Theorem 4), and in particular by (ab), which gives a contradiction.

3. Example of a graph not having any non-identical group.

Definition. We shall describe a graph G by the symbol

$$G\equiv[rs,rt,\cdots,xy,\cdots],$$

where the couples $rs,rt,\cdots,xy,\cdots$ indicate that G contains the arcs $\widehat{rs},\widehat{rt},\cdots$ etc.

By the corollaries to Theorem 3, and by Theorem 4, we have shown the

existence of groups that do not have any graphs. We now give an example of a graph that does not have any non-identical group. Let M be the graph

$$M \equiv [qr, qs, qt, qv, qw, ru, rv, rw, st, su, tw, uv].$$

Let τ be any map of M into itself. We shall show that τ is the identity. Since r is the only vertex of degree 4, $\tau(r) = r$, (Theorem 1). Since q is the only vertex of degree 5, τ also leaves q fixed, and hence carries the arc $\widehat{qr}$ into itself and permutes the vertices s, t, v and w adjacent to q among themselves. Since $I_v{}^r = 1$, $I_s{}^r = 0$, τ cannot carry v into s. Since $I_v{}^r = 1, I_t{}^r = 0$, $\tau(v) \neq t$. Suppose that $\tau(v) = w$, then τ carries the set of vertices (q, r, u) adjacent to v into the set (q, r, t) adjacent to w. Since q and r are fixed, it follows that $\tau(u) = t$. But since $I_u{}^r = 1$, $I_t{}^r = 0$, τ cannot carry u into t. Hence $\tau(v) \neq w$, and the only remaining possibility is $\tau(v) = v$. Since τ leaves v fixed it permutes the set of vertices (q, r, u) adjacent to v among themselves, and since q and r are fixed it follows that u also is fixed. Since u is fixed τ permutes the vertices (r, v, s) adjacent to u, and since r and v are fixed, s must also be fixed. Since s is fixed τ permutes the pair of vertices (u, t) adjacent to s, and since u is fixed, it follows that t is also fixed. Hence w, the only remaining vertex is also fixed. Thus we have shown that τ leaves every vertex of M fixed, hence also the arcs of M. That is M is invariant under τ and $\tau = 1$.

We note that if a graph of Degree α_0 does not have a non-identical group of degree α_0, it does not necessarily follow that this graph does not have any non-identical group whatsoever. For example, the graph G_3 of Sec. II below is of Degree 5, and does not have any non-identical group of degree 5, but has the octic group (of degree 4).

4. Preliminary Lemmas. The proofs of many theorems of Section III are based on the following lemmas, which apply to graphs of any Degree.

LEMMA 1. *If a graph is mapped into itself by a transitive group of substitutions on n of its vertices $a_1, \cdots, a_n$, then these vertices are all of the same degree.*

Proof. Suppose that a_i and a_j are of different degrees. Let σ be the substitution which carries a_i into a_j. Then vertices adjacent to a_i are carried by σ into vertices adjacent to a_j, each arc $\widehat{a_ip}$ incident with a_i being carried into an arc $\widehat{a_jq}$ incident with a_j, and each arc $\widehat{a_jq'}$ incident with a_j being the image of an arc incident with a_i. Since σ is one-one, this is impossible if a_i and a_j are of different degrees.

COROLLARY. *The group* $\mathfrak{G} \equiv (abcd)_4$ *does not have any graph.*

Proof. Suppose that $\mathfrak{G}$ has the graph G. $\mathfrak{G}$ consists of the substitutions 1, $(ab)(cd)$, $(ac)(bd)$ and $(ad)(bc)$. Since $\mathfrak{G}$ is transitive, a, b, c, d must all have the same degree in G. Also if p is any vertex of G distinct from a, b, c, d, then every substitution of $\mathfrak{G}$ leaves p fixed and $I_a{}^p = I_b{}^p = I_c{}^p = I_d{}^p$. Let $K = G$ — [the set of arcs of G not joining pairs of the vertices a, b, c, d]. Then in K, a, b, c, d are also all of the same degree.

If a, b, c, d are each of degree 1 in K, then K consists of the two arcs $\widehat{ab}, \widehat{cd}$ or of the two arcs $\widehat{ac}, \widehat{bd}$, or of the two arcs $\widehat{ad}, \widehat{bc}$. But then G is mapped into itself by the substitutions (ab), (ac) or (ad) respectively. Since $\mathfrak{G}$ contains neither of these substitutions, this gives a contradiction.

If a, b, c, d are each of degree 2 in K, then K consists of a circuit, say $\widehat{abcd}$. But then G is mapped into itself by the substitution $(abcd)$, which is not contained in $\mathfrak{G}$, giving a contradiction.

Finally, if a, b, c, d are each of degree 3 in K, then K is a complete 4-point. But then G is mapped into itself by the substitution $(abcd)$, again giving a contradiction.

LEMMA 2. *If a graph G is mapped into itself by the substitution $\sigma = (abc)(def)$ on six of its vertices a, b, c, d, e, f, then G is mapped into itself by every substitution of some one of the groups* $\mathfrak{G} \equiv (abc \cdot def)$all, $\mathfrak{G}' \equiv (abc \cdot fde)$all, *or* $\mathfrak{G}'' \equiv (abc \cdot efd)$all.[5]

Proof. σ leaves fixed any vertex p distinct from a, b, c, d, e, f. Hence

$$I_a{}^p = I_b{}^p = I_c{}^p \text{ and } I_d{}^p = I_e{}^p = I_f{}^p. \tag{3}$$

Also $I_d{}^a = I_e{}^b = I_f{}^c \equiv I_1$, $I_d{}^b = I_e{}^c = I_f{}^a \equiv I_2$, $I_d{}^c = I_e{}^c = I_f{}^b \equiv I_3$ and

$$I_b{}^a = I_c{}^b = I_a{}^c, \qquad I_e{}^d = I_f{}^e = I_d{}^f. \tag{4}$$

Case 1.

$$I_2 = I_3 \neq I_1 \tag{5}$$

Consider the substitution $\tau = (ab)(de)$. If pq is any arc not incident with a, b, d, or e, then τ leaves p, q, and $\widehat{pq}$ fixed and hence $I_q{}^p = I_{\tau(q)}{}^{\tau(p)}$. From this and examination of relations (3), (4) and (5) we see that for any pair of

[5] The substitutions of $\mathfrak{G}$ are given by G. A. Miller, "Memoir on the Substitution Groups whose degree does not exceed eight," *American Journal of Mathematics*, vol. 21 (1899), pp. 287-337, $\mathfrak{G}' \equiv [1, (abc)(def), (acb)(dfe), (ab)(df), (bc)(de), (ac)(ef)]$, $\mathfrak{G}'' \equiv [1, (abc)(def), (acb)(dfe), (ab)(ef), (bc)(fd), (ac)(de)]$.

vertices, u, v of G, $I_v^u = I_{\tau(v)}^{\tau(u)}$. Hence by Theorem 1, τ maps G into itself. Since $\mathfrak{G}$ can be generated by σ and τ, any substitutes of $\mathfrak{G}$ maps G into itself.

Case 2. $I_1 = I_3 \neq I_2$. By analogy with case 1 we can show that G is mapped into itself by the substitution $\nu = (ac)(ef)$. Since $\mathfrak{G}'$ can be generated by σ and ν, any substitution of $\mathfrak{G}'$ maps G into itself.

Case 3. $I_1 = I_2 \neq I_3$. By analogy with case 1 we can show that G is mapped into itself by the substitution $\rho = (bc)(df)$. Since $\mathfrak{G}''$ can be generated by σ and ρ, any substitution of $\mathfrak{G}''$ maps G into itself.

Case 4. $I_1 = I_2 = I_3$. As in case 1, we can show that G is mapped into itself by the substitution τ, and that any substitution of $\mathfrak{G}$ maps G into itself.

LEMMA 3. *If a graph G is mapped into itself by the substitution $\tau = (abcd)(ef)$ on six of its vertices a, b, c, d, e, f, then G is mapped into itself by every substitution of the group* $\mathfrak{G} \equiv \{(abcd)_8\text{com}(ef)\}\text{dim}$.

Proof. Suppose that G is mapped into itself by τ. If q is a vertex distinct from $a, \cdots, f$, then $I_a^q = I_b^q = I_c^q = I_d^q$ and $I_e^q = I_f^q$. Also $I_b^a = I_c^b = I_d^c = I_a^d$, $I_e^a = I_e^c$, $I_f^a = I_f^c$, $I_f^b = I_e^c$, $I_e^c = I_f^d$, $I_f^d = I_e^a$.

Consider the substitution $\sigma = (ac)$. If $\widehat{pq}$ is an arc not incident with a or c then σ leaves p, q, and $\widehat{pq}$ fixed, hence $I_q^p = I_{\sigma(q)}^{\sigma(p)}$. If r is any vertex distinct from a or c, from the preceding paragraph we see that $I_a^r = I_c^r = I_{\sigma(a)}^{\sigma(r)}$ and $I_c^r = I_{\sigma(c)}^{\sigma(r)}$. Also $I_c^a = I_a^c$, hence $I_v^u = I_{\sigma(v)}^{\sigma(u)}$ for any pair of vertices u and v. That is, σ maps G into itself (Theorem 1). Since $\mathfrak{G}$ is generated by σ and τ, every substitution of $\mathfrak{G}$ maps G into itself.

LEMMA 4. *If a graph G is mapped into itself by the substitutions $\tau = (ab)(cd)$ and $\sigma = (cd)(ef)$ on six of its vertices a, b, c, d, e, f, then G is mapped into itself by every substitution of the group* $\mathfrak{G} \equiv (ab)(cd)(ef)$.

Proof. G is mapped into itself by τ. Let q be a vertex distinct from a, b, c, d. Then $I_a^q = I_b^q$, $I_c^q = I_d^q$, and $I_c^a = I_d^b$, $I_c^b = I_d^a$.

G is also mapped into itself by σ, hence $I_c^a = I_d^a$ and $I_c^b = I_d^b$. Combining these adjacency relations with those of the preceding paragraph, we have $I_c^a = I_d^a = I_c^b = I_d^b$.

Consider the substitution $\mu = (ab)$. By discussion similar to that in the proof of Lemma 3, we can show that μ maps G into itself. Since $\mathfrak{G}$ can be generated by τ, σ, and μ, every substitution of $\mathfrak{G}$ maps G into itself.

II. Groups of the graphs of Degree ≤ 6.[9]

The proofs of the theorems in this section, and Section III, are repetitious in form, therefore to keep the length of this paper within reasonable bounds we shall for the most part delete these proofs, retaining only a few to illustrate the methods used, or those that differ materially from the general method. Except where otherwise stated, the omitted proofs are similar in form to those of Theorems 2. 1 and/or 2. 2 below.

Definition. We shall say that a graph is *admissible* if it is connected and contains no simple loops and no vertex of degree < 3.

1. Graphs of Degree 4. The only admissible graph is the complete 4-point N_4. N_4 has the group $\mathfrak{S}_4 \equiv (abcd)$all, (Theorem 3).

2. Graphs of Degree 5. There are three admissible graphs,

$G_1 \equiv N_5$, the complete 5-point.
$G_2 \equiv [ab, ac, ad, ae, bc, bd, be, cd, ce]$
$G_3 \equiv [ac, ad, ae, bc, bd, be, ce, de]$.

The graph G_1 has the group $\mathfrak{G}_8 \equiv (abcde)$all, (Theorem 3).

THEOREM 2. 1. *G_2 has the group $\mathfrak{G}_5 \equiv (abc)all(de)$.*

Proof. It is readily verified that any substitution of $\mathfrak{G}_5$ maps G_2 into itself. Suppose that G_2 is mapped into itself by a substitution τ. Since d and e are the only vertices of degree 3, τ either leaves this pair fixed or interchanges them.

Case 1. τ leaves d and e fixed. Then $\tau = \binom{abc}{\alpha\beta\gamma}$, where α, β, γ is some permutation of the letters a, b, c. Now $\mathfrak{G}_5$ contains a substitution σ such that $\sigma^{-1} = \binom{\alpha\beta\gamma}{abc}$. Hence $\tau\sigma^{-1} = 1$ and $\tau = \sigma \varepsilon \mathfrak{G}_5$.

Case 2. τ interchanges d and e. Then $\tau = \binom{abc}{\alpha\beta\gamma}(de)$. Since $\mathfrak{G}_5$ contains the substitution $\mu = (de)$, $\tau\sigma^{-1}\mu = 1$. Hence $\tau = \mu\sigma \varepsilon \mathfrak{G}_5$.

THEOREM 2. 2. *G_3 has the octic group on the four letters a, b, c, d. G_3 does not have a group of degree 5.*

Proof. Since G_3 has only one vertex e of degree 4, it cannot have a group of degree 5 (Theorem 1).

It can readily be verified that G_3 is mapped into itself by the substitution $\sigma = (abcd)$. Hence by Theorem 4, G_3 is mapped into itself by every substitution of the octic group $\mathfrak{N}_4$ $(\equiv \mathfrak{F}_5)$.

Suppose that G_3 is mapped into itself by the substitution τ. Since e is the only vertex of degree 4, τ leaves e fixed.

Case 1. $\tau(a) = a$. Then since $I_c^a = 1$, $I_d^a = 1$, either $\tau(c) = c$, $\tau(d) = d$, in which case $\tau(b) = b$, $\tau = 1$, or else $\tau(c) = d$, $\tau(d) = c$, in which case $\tau(b) = b$ and $\tau = (cd) \varepsilon \mathfrak{N}_4$.

Case 2. $\tau(a) = b$. Then since $I_c^a = I_d^a = I_c^b = 1$, either $\tau(c) = d$, $\tau(d) = c$, and $\tau = (ab)(cd) \varepsilon \mathfrak{N}_4$, or $\tau(c) = c$, $\tau(d) = d$ and $\tau = (ab) \varepsilon \mathfrak{N}_4$.

Case 3. $\tau(a) = c$ or d. A similar discussion shows that $\tau \varepsilon \mathfrak{N}_4$.

3. Graphs of Degree 6. There are eighteen admissible graphs,

$H_1 \equiv N_6$, the complete 6-point.
$H_2 \equiv [ab, ac, ad, ae, af, bc, bd, be, bf, cd, ce, cf, de, df]$
$H_3 \equiv [ab, ac, ad, ae, af, bc, bd, be, bf, cd, ce, cf, de]$
$H_4 \equiv [ac, ad, ae, af, bc, bd, be, bf, ce, cf, de, df, ef]$
$H_5 \equiv [ac, ad, ae, af, bc, bd, be, bf, ce, cf, de, df]$
$H_6 \equiv [ac, ad, ae, af, bc, bd, be, bf, cd, cf, df, ef]$
$H_7 \equiv [ab, ac, ae, af, bd, be, bf, ce, cf, de, df, ef]$
$H_8 \equiv [ab, ad, ae, af, bc, be, bf, ce, cf, de, df]$
$H_9 \equiv [ab, ac, ae, af, bd, be, bf, cd, cf, df, ef]$
$H_{10} \equiv [ad, ae, af, bd, be, bf, cd, ce, cf, df, ef]$
$H_{11} \equiv [ab, ac, ad, ae, bc, bd, bf, cd, ce, df, ef]$
$H_{12} \equiv [ad, ae, af, bc, be, bf, ce, cf, de, df, ef]$
$H_{13} \equiv [ad, ae, af, bc, be, bf, ce, cf, de, df]$
$H_{14} \equiv [ad, ae, af, bc, be, bf, cd, ce, df, ef]$
$H_{15} \equiv [ab, ae, af, bc, bf, cd, cf, de, df, ef]$
$H_{16} \equiv [ad, ae, af, bd, be, bf, cd, ce, cf, de]$
$H_{17} \equiv [ac, ad, ae, bd, be, bf, ce, cf, df]$
$H_{18} \equiv [ad, ae, af, bd, be, bf, cd, ce, cf]$

By Theorem 3, the graph H_1 has the group $\mathfrak{H}_{37} \equiv (abcdef)$ all.

THEOREM 3.1. *H_2 has the group* $\mathfrak{H}_{31} \equiv (abcd)\text{all}(ef)$.

THEOREM 3.2. *H_3 has the group* $\mathfrak{G}_5 \equiv (abc)\text{all}(de)$ *on five of its vertices a, b, c, d, e. H_3 does not have a group of degree* 6.

THEOREM 3.3. *H_4 has the group* $\mathfrak{H}_{19} \equiv (abcd)_8(ef)$.

THEOREM 3.4. *H_5 has the group* $\mathfrak{H}_{19} \equiv (abcd)_8(ef)$.

H_5 is homeomorphic with the graph formed by the edges and vertices of the regular octahedron.

THEOREM 3.5. *H_6 has the group $\mathfrak{F}_2 \equiv (ab)(cd)$, on four letters. H_6 does not have a group of degree 6.*[6]

THEOREM 3.6. *H_7 has the group $\mathfrak{H}_3 \equiv (ab \cdot cd)(ef)$.*

THEOREM 3.7. *H_8 has the group $\mathfrak{H}_3 \equiv (ab \cdot cd)(ef)$.*

THEOREM 3.8. *H_9 has the group $\mathfrak{F}_1 \equiv (ab \cdot cd)$ on four letters. H_9 does not have any group of degree 6.*[7]

THEOREM 3.9. *H_{10} has the group $\mathfrak{G}_5 \equiv (abc)\mathrm{all}(de)$, on five letters. H_{10} does not have a group of degree 6.*[8]

THEOREM 3.10. *H_{11} has the group $\mathfrak{H}_{13} \equiv \{(abcd)_8\mathrm{com}(ef)\}\mathrm{dim}$.*

THEOREM 3.11. *H_{12} has the group $\mathfrak{H}_{19} \equiv (abcd)_8(ef)$.*

THEOREM 3.12. *H_{13} has the group $\mathfrak{H}_{19} \equiv (abcd)_8(ef)$.*

THEOREM 3.13. *H_{14} has the group $\mathfrak{H}_1 \equiv (ab \cdot cd \cdot ef)$.*

THEOREM 3.14. *H_{15} has the group $\mathfrak{H}_{12} \equiv (abcde)_{10}$ on five letters. H_{15} does not have a group of degree 6.*

Proof. Since H_{15} has but one vertex f of degree 5, by Theorem 2 its group cannot be of degree 6. Since f is adjacent to each of the other vertices, any map of K into itself, where $K \equiv H_{15} - [f +$ the set of arcs incident with $f]$, is also a map of H_{15} into itself. That is, the group of K is also the group of H_{15}. Since K is the circuit $\widehat{abcde}$, K has the group $(abcde)_{10}$.

THEOREM 3.15. *H_{16} has the group $\mathfrak{G}_5 \equiv (abc)\mathrm{all}(de)$, on five letters. H_{16} does not have a group of degree 6.*

THEOREM 3.16. *H_{17} has the group $\mathfrak{H}_{17} \equiv (abcdef)_{12}$.*

Proof. It is readily verified that H_{17} is mapped into itself by every substitution of $\mathfrak{H}_{17}$. Suppose that H_{17} is mapped into itself by a non-identical substitution τ. Then some vertex of H_{17} is not left fixed by τ, say $\tau(a) \neq a$. Let $\alpha = \tau(a)$. Since $\mathfrak{H}_{17}$ is transitive it contains a substitution σ such that

[6] The group $\mathfrak{F}_2$ is simply isomorphic with $\mathfrak{H}_3 \equiv (ab \cdot cd)(ef)$ of degree 6; *cf.* Miller, *loc. cit.*

[7] $\mathfrak{F}_1$ is simply isomorphic with $\mathfrak{H}_1 \equiv (ab \cdot cd \cdot ef)$ of degree 6, Miller, *loc. cit.*

[8] $\mathfrak{G}_5$ is simply isomorphic with $\mathfrak{H}_{17} \equiv (abcdef)_{12}$ of degree 6, *cf.* Miller, *loc. cit.*

$\sigma(\alpha) = a$. Then $\tau' = \tau\sigma$ maps H_{17} into itself and leaves a fixed. Since $I_c^a = 1$, $I_b^a = 0$, $I_f^a = 0$ τ' cannot carry c into b or f. Similarly τ' cannot carry d or e into b or f. That is, τ' permutes the vertices c, d, e among themselves, and hence either leaves the pair b, f fixed or interchanges them.

Case 1. τ' leaves fixed the pair b, f. Since $I_d^f = 1$, $I_e^f = 0$, $\tau'(d) \neq e$. Since $I_d^b = 1$, $I_c^b = 0$, $\tau'(d) \neq c$. Hence τ' leaves d fixed. Since $I_c^f = 1$, $I_e^f = 0$, $\tau'(c) \neq e$. Hence τ' leaves c fixed, and therefore the remaining vertex e is left fixed. That is $\tau' = \tau\sigma = 1$. Hence $\tau = \sigma^{-1} \varepsilon \mathfrak{H}_{17}$.

Case 2. τ' interchanges b and f. Let $\mu = (bf)(ce) \varepsilon \mathfrak{H}_{17}$. Then $\tau'' = \tau'\mu$ maps H_{17} into itself, leaving fixed the vertices a, b, f and permuting the vertices c, d, e among themselves. As in case 1, we can show that $\tau'' = \tau\sigma\mu = 1$. That is, $\tau = \mu^{-1}\sigma^{-1} \varepsilon \mathfrak{H}_{17}$.

THEOREM 3.17. *H_{18} has the group $\mathfrak{H}_{34} \equiv (abcdef)_{72}$.*

The proof is similar to that of Theorem 3.16.

The graph H_{18} is of special interest, since it is one of the two irreducible non-planar graphs (the other being the complete 5-point).

Since we have shown the existence of a non-identical group for each admissible graph of Degree ≤ 6, we have proved the following theorem:

THEOREM 6. *Any graph of Degree ≤ 6, containing no simple loops, and no vertices of degree < 3, has a non-identical group.*

III. Graphs of the substitution groups of degree ≤ 6.

A. Groups of degree 2 and 3.[9] The groups are,

$$\mathfrak{E}_1 \equiv (ab), \qquad \mathfrak{E}_2 \equiv (abc)\text{cyc}, \qquad \mathfrak{E}_3 \equiv (abc)\text{all}.$$

THEOREM A.1. *The group $\mathfrak{E}_1$ has the graph*

$$E_1 \equiv [ab] + M + [\textit{the set of arcs joining } a \textit{ and } b \textit{ to each vertex of } M],$$

where M is the graph of Section I, 3.

---

[9] A complete list of the substitutions of $\mathfrak{H}_{29}$ and $\mathfrak{H}_{32}$ are given by Cole, "List of substitution groups of Nine Letters," *Quarterly Journal of Mathematics*, vol. 26 (1892), p. 372 *et seq.*

A complete list of the substitutions of $\mathfrak{H}_{18}$, $\mathfrak{H}_{24}$, $\mathfrak{H}_{27}$, $\mathfrak{H}_{28}$, $\mathfrak{H}_{31}$, $\mathfrak{H}_{33}$, $\mathfrak{H}_{34}$, $\mathfrak{H}_{35}$, are given by Cayley, "On substitution groups . . . etc.," *Quarterly Jounral of Mathematics*, vol. 25 (1891), pp. 71-88.

A complete list of the substitutions of all the other groups are given by Miller, *loc. cit.*

The proof is similar in form to that of Theorem B. 1 below.

The group $\mathfrak{E}_2$ does not have any graph ((Theorem 4, Corollary 1).

THEOREM A. 2. *The group* $\mathfrak{E}_3$ *has the graph*

$E_3 \equiv [ab, ac, bc] + M +$ [*the set of arcs joining a, b, c to each vertex of M*], *where M is the graph of Section* I, 3.

The proof is similar in form to that of the first paragraph of the proof of Theorem B. 1 and the proof of Theorem 2. 2.

B. Groups of degree 4.[9] There are seven groups of degree 4. They are, in sequence of increasing order,

$$\mathfrak{F}_1 \equiv (ab \cdot cd), \quad \mathfrak{F}_2 \equiv (ab)(cd), \quad \mathfrak{F}_3 \equiv (abcd)_4, \quad \mathfrak{F}_4 \equiv (abcd)\,\mathrm{cyc},$$
$$\mathfrak{F}_5 \equiv (abcd)_8, \quad \mathfrak{F}_6 \equiv (abcd)\,\mathrm{pos}, \quad \mathfrak{F}_7 \equiv (abcd)\,\mathrm{all}.$$

The group $\mathfrak{F}_1$ has the graph H_9, (Theorem 3. 8).

THEOREM B. 1. *The group* $\mathfrak{F}_2$ *has the graph*

$$F_2 \equiv [ac, ad, bc, bd, cd] + M$$

+ [*the set of arcs joining each of the vertices a, b, c, d to each vertex of M*] *where M is the graph of Section* I, 3.

Proof. It is readily verified that the substitutions of $\mathfrak{F}_2$ map F_2 into itself. Suppose that F_2 is mapped into itself by a substitution τ. Since c and d are the only vertices of degree 10, τ either leaves this pair fixed or interchanges them. Since r is the only vertex of degree 8, τ leaves r fixed. By a proof similar to that used to show that M has no non-identical group, we can show that τ leaves each of the vertices s, t, u, v, w fixed. The vertices a, b, q are each of degree 9. Since $I_u{}^a = 1$, $I_u{}^q = 0$, τ cannot carry a into q. Similarly $\tau(b) \neq q$. Hence τ either leaves the pair a, b fixed or interchanges them. Consequently q is left fixed. That is τ leaves every vertex of M fixed, and thus also every arc of M.

Case 1. τ leaves a and b fixed. Then if τ leaves c and d fixed, $\tau = 1$; while if τ interchanges c and d, then $\tau(\widehat{cq}) = \widehat{dq}, \cdots, \tau(\widehat{cw}) = \widehat{dw}$, $\tau(\widehat{dq}) = \widehat{cq}, \cdots, \tau(\widehat{dw}) = \widehat{cw}$. That is $\tau = (cd) \,\varepsilon\, \mathfrak{F}_2$.

Case 2. τ interchanges a and b. Then if τ leaves c and d fixed, $\tau = (ab) \,\varepsilon\, \mathfrak{F}_2$; while if τ interchanges c and d, $\tau = (ab)(cd) \,\varepsilon\, \mathfrak{F}_2$.

The group $\mathfrak{F}_3$ does not have any graph (Lemma 1, Corollary).
The group $\mathfrak{F}_4$ does not have any graph (Theorem 4, Corollary 1).
The group $\mathfrak{F}_5$ has the graph G_3 (Theorem 2. 2).
The group $\mathfrak{F}_6$ does not have any graph (Theorem 5).
The group $\mathfrak{F}_7$ has the complete 4-point for its graph (Theorem 3).

C. Groups of degree 5.[9] There are eight groups of degree 5. They are, in sequence of increasing order,

$$\mathfrak{G}_1 \equiv (abcde)\text{cyc}, \quad \mathfrak{G}_2 \equiv (abc)\text{cyc}(de), \quad \mathfrak{G}_3 \equiv \{(abc)\text{all}(de)\}\text{pos}$$
$$\mathfrak{G}_4 \equiv (abcde)_{10}, \quad \mathfrak{G}_5 \equiv (abc)\text{all}(de), \quad \mathfrak{G}_6 \equiv (abcde)_{20},$$
$$\mathfrak{G}_7 \equiv (abcde)\text{pos}, \quad \mathfrak{G}_8 \equiv (abcde)\text{all}.$$

The group $\mathfrak{G}_1$ does not have any graph (Theorem 4, Corollary 1).
The group $\mathfrak{G}_2$ does not have any graph (Theorem 4, Corollary 2).

THEOREM C. 1. *$\mathfrak{G}_3$ does not have any graph.*

Proof. Suppose that $\mathfrak{G}_3$ has a graph G. Then G is mapped into itself by the substitution (abc) on three of its vertices a, b, c. But then by Theorem 4, G is mapped into itself by the substitution (ab). Since G_3 does not contain (ab) this gives a contradiction.

THEOREM C. 2. *$\mathfrak{G}_4$ has the graph*

$$G_4 \equiv [ab, ae, bc, cd, de] + M$$

+ [the set of arcs joining each of the vertices a, b, c, d, e to each vertex of M], where M is the graph of Section I, 3.

The proof is similar to that of Theorem B. 1.

The group $\mathfrak{G}_5$ has the graph G_2 (Theorem 2. 1), and also the graphs H_3, H_{10} and H_{16} (Theorems 3. 2, 3. 9, 3. 15).

THEOREM C. 3. *$\mathfrak{G}_6$ does not have any graph.*

The proof is similar to that of Theorem C. 1.

The group $\mathfrak{G}_7$ does not have any graph (Theorem 5).
The group $\mathfrak{G}_8$ has the complete 5-point for its graph (Theorem 3).

D. Groups of degree 6.[9] There are thirty-seven groups of degree 6. Arranged in sequence of increasing order they are,

$\mathfrak{H}_1 \equiv (ab \cdot cd \cdot ef)$; $\mathfrak{H}_2 \equiv (abc \cdot def)\mathrm{cyc}$; $\mathfrak{H}_3 \equiv (ab \cdot cd)(ef)$
$\mathfrak{H}_4 \equiv \{(ab)(cd)(ef)\}\mathrm{pos}$; $\mathfrak{H}_5 \equiv \{(abcd)_4(ef)\}\mathrm{dim}$; $\mathfrak{H}_6 \equiv \{(abcd)\mathrm{cyc}(ef)\}\mathrm{pos}$;
$\mathfrak{H}_7 \equiv (abcdef)\mathrm{cyc}$; $\mathfrak{H}_8 \equiv (abc \cdot def)\mathrm{all}$; $\mathfrak{H}_9 \equiv (abcdef)_6$;
$\mathfrak{H}_{10} \equiv (ab)(cd)(ef)$; $\mathfrak{H}_{11} \equiv (abcd)\mathrm{cyc}(ef)$; $\mathfrak{H}_{12} \equiv (abcd)_4(ef)$
$\mathfrak{H}_{13} \equiv \{(abcd)_8\mathrm{com}(ef)\}\mathrm{dim}$; $\mathfrak{H}_{14} \equiv \{(abcd)_8\mathrm{cyc}(ef)\}\mathrm{dim}$; $\mathfrak{H}_{15} \equiv \{(abcd)_8\mathrm{pos}(ef)\}\mathrm{dim}$;
$\mathfrak{H}_{16} \equiv (abc)\mathrm{cyc}(def)\mathrm{cyc}$; $\mathfrak{H}_{17} \equiv (abcdef)_{12}$; $\mathfrak{H}_{18} \equiv (abcdef)_{12_2}$;
$\mathfrak{H}_{19} \equiv (abcd)_8(ef)$; $\mathfrak{H}_{20} \equiv (abc)\mathrm{all}(def)\mathrm{cyc}$; $\mathfrak{H}_{21} \equiv \{(abc)\mathrm{all}(def)\mathrm{all}\}\mathrm{pos}$;
$\mathfrak{H}_{22} \equiv (abcdef)_{18}$; $\mathfrak{H}_{23} \equiv (abcd)\mathrm{pos}(ef)$; $\mathfrak{H}_{24} \equiv \{(abcd)\mathrm{all}(ef)\}\mathrm{pos}$;
$\mathfrak{H}_{25} \equiv (\pm\, abcdef)_{24}$; $\mathfrak{H}_{26} \equiv (+\, abcdef)_{24}$; $\mathfrak{H}_{27} \equiv (abcdef)_{24_5}$;
$\mathfrak{H}_{28} \equiv (abc)\mathrm{all}(def)\mathrm{all}$; $\mathfrak{H}_{29} \equiv (abcdef)_{36}$; $\mathfrak{H}_{30} \equiv (abcdef)_{36_3}$;
$\mathfrak{H}_{31} \equiv (abcd)\mathrm{all}(ef)$; $\mathfrak{H}_{32} \equiv (abcdef)_{48}$; $\mathfrak{H}_{33} \equiv (abcdef)_{60}$;
$\mathfrak{H}_{34} \equiv (abcdef)_{72}$; $\mathfrak{H}_{35} \equiv (abcdef)_{120}$; $\mathfrak{H}_{36} \equiv (abcdef)\mathrm{pos}$;
$\mathfrak{H}_{37} \equiv (abcdef)\mathrm{all}$.

The group $\mathfrak{H}_1$ has the graph H_{14} (Theorem 3. 13).

THEOREM D. 1. *The groups $\mathfrak{H}_2$ and $\mathfrak{H}_9$ have no graphs.*

The proofs, based on Lemma 2, are similar to that of Theorem C. 1.

The group $\mathfrak{H}_3$ has the graph H_7 (Theorem 3. 6).

THEOREM D. 2. *The groups $\mathfrak{H}_4$ and $\mathfrak{H}_{18}$ have no graphs.*

The proofs, based on Lemma 4, are similar to that of Theorem C. 1.

THEOREM D. 3. *The group $\mathfrak{H}_5$ has the graph*

$$R \equiv [ab, ac, ae, ag, bd, be, bg, cd, cf, cg, df, dg, eg, fg].$$

THEOREM D. 4. *The groups $\mathfrak{H}_6$, $\mathfrak{H}_{15}$, and $\mathfrak{H}_{26}$ have no graphs.*

The proofs, based on Lemma 3, are similar to that of Theorem C. 1.

The group $\mathfrak{H}_7$ does not have any graph (Theorem 4, Corollary 1).

THEOREM D. 5. *The group $\mathfrak{H}_8$ has the graph*

$$L \equiv [ab, ac, ae, af, ag, bc, bd, bf, bg, cd, ce, cg, dg, eg, fg].$$

THEOREM D. 6. *The group $\mathfrak{H}_{10}$ has the graph*

$$N \equiv [ac, ad, ag, bc, bd, bg, ce, cf, cg, de, df, dg, ef].$$

THEOREM D. 7. *The groups $\mathfrak{H}_{11}$, $\mathfrak{H}_{14}$, $\mathfrak{H}_{25}$, and $\mathfrak{H}_{35}$ have no graphs.*

The proofs, based on Theorem 4, are similar to that of Theorem C. 1.

THEOREM D. 8. *$\mathfrak{H}_{12}$ does not have any graph.*

Proof. $\mathfrak{H}_{12}$ contains as subgroup the group $\mathfrak{F}_3 \equiv (abcd)_4$. Suppose that $\mathfrak{H}_{12}$ has a graph H. Then H is mapped into itself by all the substitutions of $\mathfrak{F}_3$.

But then, as is shown in the proof of the corollary to Lemma 1, H is also mapped into itself by one of the substitutions (ab), (ac), (ad), $(abcd)$, $(abdc)$, $(acbd)$, $(acdb)$, $(adbc)$ or $(adcb)$. Since $\mathfrak{H}_{12}$ contains neither of these substitutions, this gives a contradiction.

The group $\mathfrak{H}_{13}$ has the graph H_{11} (Theorem 3.10).

THEOREM D. 9. *The groups* $\mathfrak{H}_{16}$, $\mathfrak{H}_{20}$, $\mathfrak{H}_{21}$, $\mathfrak{H}_{22}$, $\mathfrak{H}_{23}$, $\mathfrak{H}_{24}$, $\mathfrak{H}_{29}$ *and* $\mathfrak{H}_{30}$ *have no graphs.*

The proofs, based on Theorem 4, are similar to that of Theorem C. 1.

The group $\mathfrak{H}_{17}$ has the graph H_{17} (Theorem 3.16).
The group $\mathfrak{H}_{19}$ has the graph H_5 (Theorem 3.4).

THEOREM D. 10. $\mathfrak{H}_{27}$ *does not have any graph.*

Proof. Suppose that $\mathfrak{H}_{27}$ has a graph H. Then H is mapped into itself by the substitution $(aebcfd)$. But then H is mapped into itself by every substitution of the group $\mathfrak{H}^* \equiv (aebcfd)_{12}$, (Theorem 4),[10] and in particular by $(ab)(cd)$. Since $\mathfrak{H}_{27}$ does not contain the latter, this gives a contradiction.

THEOREM D. 11. *The group* $\mathfrak{H}_{28}$ *has the graph*

$$P \equiv [ab, ac, ag, bc, bg, cg, de, df, dg, dh, ef, eg, eh, fg, fh].$$

The group $\mathfrak{H}_{31}$ has the graph H_2 (Theorem 3.1).

THEOREM D. 12. $\mathfrak{H}_{32}$ *does not have any graph.*

The proof is similar to that of Theorem D. 11, involving the group $\mathfrak{H}^{**} \equiv (abecfd)_{12}$ and the particular substitution $(aef)(bcd)$.[10]

THEOREM D. 13. $\mathfrak{H}_{33}$ *does not have any graph.*

The proof, based on Theorem 4, is similar to that of Theorem C. 1.

The group $\mathfrak{H}_{34}$ has the graph H_{18} (Theorem 3.17).
The group $\mathfrak{H}_{36}$ does not have any graph (Theorem 5).
The group $\mathfrak{H}_{37}$ has for its graph the complete 6-point (Theorem 3).

NEW YORK CITY.

[10] Miller (*loc. cit.*) gives the substitutions of the group $(abcdef)_{12}$. The substitutions of $\mathfrak{H}^*$, $\mathfrak{H}^{**}$ can be obtained from these by making the obvious indicated changes in notation.

TERNARY AND QUATERNARY ELIMINATION.*

By Arthur B. Coble.

---

1. Introduction. It is the aim of the present paper to exhibit explicit determinants of covariant character which will, in certain cases, be the resultant R of n forms in n variables, $(\alpha_i y)^{l_i}$ $(i = 1, \cdots, n)$, multiplied by a relatively simple factor also of covariant character.

Van der Waerden [1] (p. 20) defines R to be the G. C. D. of $l+1$ determinants found in the "dialytic array" of forms of order $l+1$ where $l = \Sigma(l_i - 1)$. Macaulay [2] (p. 4) had earlier defined R to be the G. C. D. of all of the determinants of this array and proves (pp. 10-11) that $D = E \cdot R$, where D is a specific determinant of the array and E is a specific minor of D. This result is quite precise but neither D, nor the "extraneous factor" E are invariantive and both are determinants of high order.

The method used is an extension of that initiated by F. Morley [3] and developed more extensively by Morley and Coble [4]. In the first of these papers Morley introduced a covariant J_k which, for the general case above, has the expression,

$$J_k(x^h, y^k) = (\alpha_1 \alpha_2 \cdots \alpha_n) \Sigma \prod_{i=1}^{i=n} (\alpha_i x)^{l_i - k_i - 1} (\alpha_i y)^{k_i}, \tag{1}$$

$$h + k = l, \qquad 0 \leq k_i \leq l_i - 1,$$

where Σ is extended over all sets of positive integer values of the k_i for which also

$$k_1 + k_2 + \cdots + k_n = k. \tag{2}$$

In the second paper it is proved that, when the n equations, $(\alpha_i y)^{l_i} = 0$, have a fixed common solution x, then $J(x^h, y^k)$ is in the module determined by the n given forms. Thus, for

$$(\alpha_i x)^{l_i} = 0,$$

$$J(x^h, y^k) \equiv (\lambda_1 y)^{k - l_1} \cdot (\alpha_1 y)^{l_1} + \cdots + (\lambda_n y)^{k - l_n} \cdot (\alpha_n y)^{l_n}. \tag{3}$$

This identity (3) in y together with suitable dialytic equations enabled

---

* Received May 18, 1946.

Morley and Coble to obtain determinants $D=R$ in all of the ternary cases covered by

$$(4) \qquad l_1 \leq l_2 \leq l_3 \leq l_1 + l_2 + 1.$$

It was used later by T. W. Moore [5] to obtain a variety of forms for a particular binary resultant. Its usefulness in the quaternary domain was much more limited. There only the lowest order l_1 was arbitrary, the others being quite limited.

We propose here to use this identity more widely and to cover all cases of ternary and quaternary elimination. The new determinants D obtained are all of the form ER but the extraneous factor E is itself a covariant of the given forms, usually of very simple character.

In the course of the work we use constantly the number of coefficients in a given ternary or quaternary form. This number, a binomial coefficient, will be indicated by the more convenient notation defined by

$$[n]_k = \binom{n}{k} = {}_nC_k.$$

2. The Ternary Identities. In order to avoid subscripts we shall use forms,

$$(5) \qquad (\alpha y)^m, (\beta y)^n, (\gamma y)^p \quad [m \leq n \leq p \geq m+n+2].$$

The fundamental identity (3) then reads:

$$(\mathrm{I}) \quad J(x^h, y^k) + (\lambda y)^{k-m} \cdot (\alpha y)^m + (\mu y)^{k-n} \cdot (\beta y)^n + (\nu y)^{k-p} \cdot (\gamma y)^p \equiv 0 \text{ (in } y),$$

$$h+k=l=m+n+p-3.$$

Naturally, if any of the exponents $k-m$, $k-n$, $k-p$ are negative, the corresponding terms do not appear. We shall assume merely that

$$h < n+p, \qquad k < n+p.$$

Then, in (I), the coefficients $(\lambda y)^{k-m}$, $(\mu y)^{k-n}$, $(\nu y)^{k-p}$ are no longer necessarily unique since terms of the form $(\kappa_{12} y)^{k-m-n} \cdot (\alpha y)^m \cdot (\beta y)^n$ can appear in either λ or μ, and terms of the form $(\kappa_{13} y)^{k-m-p} \cdot (\alpha y)^m \cdot (\gamma y)^p$ can appear in either λ or ν. When $k \leq m+n$, or when $k \leq m+p$ we restore the uniqueness of the coefficients λ, μ, ν by the respective requirements:

$$(\mathrm{IIa}) \qquad (\mu y)^{k-m-n} (\mu z)^m \equiv 0 \text{ (in } y), (\alpha z)^m \neq 0;$$

$$(\mathrm{IIb}) \qquad (\nu y)^{k-m-p} (\nu z)^m \equiv 0 \text{ (in } y), (\alpha z)^m \neq 0$$

where z is any arbitrarily chosen point not on $(\alpha y)^m \neq 0$. Indeed, if λ, μ, ν are admissible multipliers and if $k \leq m + n$, $k < n + p$, the multiplier μ may be replaced by $(\mu y)^{k-n} - (\kappa_{12} y)^{k-n-m} \cdot (\alpha y)^m$, where $(\kappa_{12} y)^{k-n-m}$ is an arbitrary form. The requirement (IIa) imposes $[k - n - m + 2]_2$ linear non-homogeneous conditions on the $[k - n - m + 2]_2$ coefficients of $(\kappa_{12} y)^{k-n-m}$. The determinant of the system has elements which are linear in the coefficients of the form $(\alpha y)^m$ and of the m-th order in z. Since the determinant is obviously a covariant it must be a numerical multiple of $[(\alpha z)^m]^{[k-m-n+2]_2}$. Thus, if $(\alpha z)^m \neq 0$, $(\mu y)^{k-n}$ is uniquely defined by (IIa), and similarly $(\nu y)^{k-p}$ is uniquely defined by (IIb), whence also $(\lambda y)^{k-m}$ is uniquely determined.

To the above two types of identities we add a third type derived from equations of the dialytic array. These equations can be expressed in the form,

$$(6) \qquad (\rho x)^{h-m} \cdot (\alpha x)^m \equiv 0 \text{ (in } \rho), \ (\sigma x)^{h-n} \cdot (\beta x)^n \equiv 0 \text{ (in } \sigma),$$
$$(\tau x)^{h-p} \cdot (\gamma x)^p \equiv 0 \text{ (in } \tau).$$

If however $h \geq m + n$ or $h \geq m + p$, the equations obtained from the identities in σ or τ will overlap with those obtained from the identity in ρ. This overlapping may be avoided by the requirement that σ, τ have at z multiple points of orders $h - m - n - 1$, $h - m - p - 1$ respectively, i. e., that

$$(7) \qquad (\sigma y)^{h-m-n} (\sigma z)^m \equiv 0 \text{ (in } y), \ (\tau y)^{h-m-p} (\tau z)^m \equiv 0 \text{ (in } y).$$

This requirement defines σ, for example, in the linear system $(\sigma y)^{h-n} + (\rho_{12} y)^{h-m-n} \cdot (\alpha y)^m$, and will bar the overlapping part, $(\rho_{12} y)^{h-m-n} \cdot (\alpha y)^m$, if $(\alpha z)^m \neq 0$. We can then replace the second identity in (6) by

$$(\sigma x)^{h-n} \cdot (\beta x)^n + b (\sigma x)^{h-m-n} (\sigma z)^m \equiv 0 \text{ (in } \sigma),$$

since the first term vanishes identically due to $(\beta x)^n = 0$, and the second term vanishes due to (7). The constant b is introduced in order to cover the lack of homogeneity in the coordinates of the fixed common point x, and in the coordinates of the extraneous point. Naturally in using such an identity we must eliminate both the combinations of x to the degree h and b times the combinations of x to the degree $h - m - n$. Finally we replace the coefficients σ of the form $(\sigma y)^{h-n}$ by the $(h - n)$-th combinations of a linear form (ηy) to obtain the third type of identity:

$$\text{(IIIa)} \qquad (\eta x)^{h-m} \cdot (\alpha x)^m \equiv 0 \text{ (in } \eta);$$

(IIIb) $$(\eta x)^{h-n}\cdot(\beta x)^n + b(\eta x)^{h-m-n}(\eta z)^m \equiv 0 \ (\text{in } \eta);$$

(IIIc) $$(\eta x)^{h-p}\cdot(\gamma x)^p + c(\eta x)^{h-m-p}(\eta z)^m \equiv 0 \ (\text{in } \eta).$$

In case the terms in b or c disappear in the identities (IIIb) or (IIIc), due to $h < m + n$ or $h < m + p$, we shall refer to them as (IIIb′) or (IIIc′) respectively.

3. Determinants D for given m, n, p. With a given set of orders m, n, p restricted as in **2** (5) we have a variety of cases depending on the choice of h, k in

(8) $$h + k = l = m + n + p - 3.$$

Each case yields a determinant D which contains the factor R. We distinguish these cases by the value of S in

(9) $$h = m + S - 3, \qquad k = n + p - S$$
$$(S = 1, 2, \cdots, p + n - m + 2),$$

with the exception that, if $m = 1$, the first and last values of S are to be omitted. On the value of k there will depend the occurrence of all, part, or none, of λ, μ, ν in the identity (I) and the use of the identities (IIa), (IIb). On the value of h there will depend the need for identities of the simple type (IIIa), (IIIb′), (IIIc′), or the replacement of (IIIb′), (IIIc′) by (IIIb), (IIIc). In all of the cases our object is to obtain, from a complete use of the identities involved, just enough equations to eliminate the combinations x^h, and such of the coefficients λ, μ, ν and combinations bx^{h-m-n}, cx^{h-m-p} as occur, by using the determinant D of the equations.

The range of values of S may be divided into sets, each set contributing a particular form of D. This division is as follows:

	S	k	h	
Set I 1:	1	$n+p-1$	$m-2$	; $\lambda, \mu(\kappa_{12}), \nu(\kappa_{13})$
	2	$n+p-2$	$m-1$	;
Set I 2:	3	$n+p-3$	m	; $\lambda, \mu(\kappa_{12}), \nu(\kappa_{13}), \alpha$
. . .	. . .	. . .	. . .	. . .
	$n-m$	$p+m$	$n-3$	;
Set II 1:	$n-m+1$	$p+m-1$	$n-2$	; $\lambda, \mu(\kappa_{12}), \nu, \alpha$
	$n-m+2$	$p+m-2$	$n-1$	;
Set II 2:	$n-m+3$	$p+m-3$	n	; $\lambda, \mu(\kappa_{12}), \nu, \alpha, \beta$
. . .	. . .	. . .	. . .	. . .

	n	p	$m+n-3$	;
Set III 1:	$n+1$	$p-1$	$m+n-2$	; $\lambda, \mu(\kappa_{12}), \alpha, \beta$
	$n+2$	$p-2$	$m+n-1$	;
Set III 2:	$n+3$	$p-3$	$m+n$	; $\lambda, \mu(\kappa_{12}), \alpha, \beta(b$
. . . .	. . . .	. . . .	. . . .	. . . .
	$p-m$	$m+n$	$p-3$	;
Set IV 1:	$p-m+1$	$m+n-1$	$p-2$	; $\lambda, \mu, \alpha, \beta(b)$
	$p-m+2$	$m+n-2$	$p-1$	;
Set IV 2:	$p-m+3$	$m+n-3$	p	; $\lambda, \mu, \alpha, \beta(b), \gamma$
. . . .	. . . .	. . . .	. . . .	. . . .
	p	n	$p+m-3$	;
Set V 1:	$p+1$	$n-1$	$p+m-2$	; $\lambda, \alpha, \beta(b), \gamma$
	$p+2$	$n-2$	$p+m-1$	;
Set V 2:	$p+3$	$n-3$	$p+m$	; $\lambda, \alpha, \beta(b), \gamma(c)$
. . . .	. . . .	. . . .	. . . .	. . . .
	$p+n-m$	m	$p+n-3$	;
Set VI 1:	$p+n-m+1$	$m-1$	$p+n-2$	; $\alpha, \beta(b), \gamma(c)$
	$p+n-m+2$	$m-2$	$p+m-1.$	

The quantities at the end of a particular set need some explanation. The occurrence of μ indicates that the coefficients μ can be eliminated by using the identity (I) alone, that of $\mu(\kappa_{12})$ indicates that identities (I) and (IIa) must both be used to eliminate the coefficients μ. The occurrence of β indicates that the dialytic identity (IIIb′) is used whereas that of $\beta(b)$ indicates that (IIIb) must be employed. Thus a glance down the last column shows that the expression for D is different in each set. However the main differences appear as we pass through Sets I 2, II 2, III 2, IV 2, V 2 and we indicate below the determinants D for these five cases. A trifling omission in any one of these five will cover either the preceding or the following set in each of which S has only two values.

The determinant D for Set I 2 has the following expression and value:

Set I 2: $D = R \cdot [(\alpha z)^m]^d =$

	$[m+S-1]_2 x^h$	$[n+p-m-S+2]_2 \lambda$	$[p-S+2]_2 \mu$	$[n-S+2]_2 \nu$
(I) : $[n+p-S+2]_2$	$\alpha\beta\gamma$	α	β	γ
(IIa) : $[p-S-m+2]_2$	0	0	z^m	0
(IIb) : $[n-S-m+2]_2$	0	0	0	z^m
(IIIa) : $[S-1]_2$	α	0	0	0

$$(d = [p-S-m+2]_2 + [n-S-m+2]_2).$$

The binomial coefficients at the left of the array give the number of equations obtained from the identities indicated; those at the top of the array give the number of terms eliminated. For the blocks of elements within the array, the degrees in the coefficients of the given forms and in z are indicated. The verification that the sum of the rows of D is equal to the sum of the columns will not be reproduced here. The expansion of a determinant of the form of D is simple, and it is easily verified that the degrees of D in the coefficients of the given forms and z are those of the product given above. That D actually is this product is proved in **6.**

The above D will adequately cover the Set I 1. For, if S is 1 or 2, $[S-1]_2=0$, and the identity (IIIa) will be missing. Thus the verifications mentioned above need not be repeated. It may be used to cover the Set II 1 also. For, if S is $n-m+1$ or $n-m+2$, $[n-S-m+2]_2=0$, and the identity (II b) will not appear.

With these explanations we may, without further ado, set down the typical forms of the determinants D for the Sets II 2, III 2, IV 2, V 2, for each of which the two verifications mentioned above are easily made, and for each of which simple alterations cover the adjacent cases.

Set II 2: $D=R\cdot[(\alpha z)^m]^d=$

	$[m+S-1]_2 x^h$	$[n+p-m-S+2]_2\lambda$	$[p-S+2]_2\mu$	$[n-S+2]_2\nu$
(I) : $[n+p-S-2]_2$	$\alpha\beta\gamma$	α	β	γ
(IIa) : $[p-m-S+2]_2$	0	0	z^m	0
(IIIa) : $[S-1]_2$	α	0	0	0
(IIIb′) : $[m-n+S-1]_2$	β	0	0	0

$$(d=[p-m-S+2]_2).$$

Set III 2: $D=R\cdot[(\alpha z)^m]^d=$

	$[m+S-1]_2 x^h$	$[n+p-m-S+2]_2\lambda$	$[p-S+2]_2\mu$	$[S-n-1]_2 b x^{S-n-3}$
(I) : $[n+p-S+2]_2$	$\alpha\beta\gamma$	α	β	0
(IIa) : $[p-m-S+2]_2$	0	0	z^m	0
(IIIa) : $[S-1]_2$	α	0	0	0
(IIIb) : $[m-n+S-1]_2$	β	0	0	z^m

$$(d=[S-n-1]_2+[p-m-S+2]_2).$$

Set IV 2: $D = R \cdot [(\alpha z)^m]^d =$

	$[m+S-1]_2 x^h$	$[n+p-m-S+2]_2 \lambda$	$[p-S+2]_2 \mu$	$[S-n-1]_2 b x^{S-n-3}$
$(\mathrm{I}) : [n+p-S+2]_2$	$\alpha\beta\gamma$	α	β	0
$(\mathrm{IIIa}) : [S-1]_2$	α	0	0	0
$(\mathrm{IIIb}) : [m-n+S-1]_2$	β	0	0	z^m
$(\mathrm{IIIc}') : [m+S-p-1]_2$	γ	0	0	0

$(d = [S-n-1]_2).$

Set V 2: $D = R \cdot [(\alpha z)^m]^d =$

	$[m+S-1]_2 x^h$	$[n+p-m-S+2]_2 \lambda$	$[S-n-1]_2 b x^{S-n-3}$	$[S-p-1]_2 c x^{S-n-3}$
$(\mathrm{I}) : [n+p-S+2]_2$	$\alpha\beta\gamma$	α	0	0
$(\mathrm{IIIa}) : [S-1]_2$	α	0	0	0
$(\mathrm{IIIb}) : [m-n+S-1]_2$	β	0	z^m	0
$(\mathrm{IIIc}) : [m-p+S-1]_2$	γ	0	0	z^m

$(d = [S-n-1]_2 + [S-p-1]_2).$

4. Reciprocity among the determinants D for given m, n, p. In all of the above determinants D, the identity (I) involving $J(x^h, y^k)$ has been employed. We consider now the relation between the determinant D for $J(x^h, y^k)$ and the determinant D' for $J(x^k, y^h)$. They arise respectively from values S and S' for which

(10) $$S + S' = p + n - m + 3.$$

Thus the Set III 2 is self-reciprocal, the first member of the set being paired with the last, etc. The Sets III 1 and IV 1 are reciprocal to each other, as also are the pairs II 2, IV 2; II 1, V 1; I 2, V 2; and I 1, VI 1. It is clear that in $J(x^h, y^k)$ and $J(x^k, y^h)$ the matrices $\alpha\beta\gamma$ are transposed. It is also clear from the last column of the table in **3** that, when h and k are interchanged, the terms in λ, μ, ν are interchanged with those in α, β, γ and that $\mu(\kappa_{12})$, $\nu(\kappa_{13})$ are interchanged with $\beta(b)$, $\gamma(c)$ respectively. From the character of the identities involved we may conclude that

(11) *If S, S' satisfy the relation* (10), *then D, D' formed for these values are conjugate determinants.*

From this there follows immediately that

(12) *The number of essentially different determinants D is*

$$(p+n-m+3)/2 \text{ or } (p+n-m+4)/2.$$

5. The non-vanishing of the determinants D. It is essential for our conclusion that D be not identically zero. This is verified by the non-vanishing of D in the particular case

$$(\alpha y)^m = y_1^m,\ (\beta y)^n = y_2^n,\ (\gamma y)^p = y_3^p;\ z_1 : z_2 : z_3 = 1 : 0 : 0, \tag{13}$$

in which case $(\alpha z)^m = 1$. We carry out the verification for only one specimen case in Set III 1, say the case

$$S = n+2, \qquad h = m+n-1, \qquad k = p-2.$$

Then

$$J(x^{m+n-1}y^{p-2}) = \Sigma x_1^{i_1}x_2^{i_2}x_3^{i_3}y_1^{m-1-i_1}\,y_2^{n-1-i_2}y_3^{p-1-i_3} \tag{14}$$
$$(i_1+i_2+i_3 = m+n-1,\ i_1 \leq m-1,\ i_2 \leq n-1).$$

The multiple-point condition on μ is now $(\mu y)^{p-m-n-2}\mu_1^m \equiv 0$. This merely asserts that $[p-m-n]_2$ coefficients of $(\mu y)^{p-n-2}$ are zero and this condition may be dropped if only the remaining coefficients of μ are retained. Thus D takes the relatively simple form

$$D = \begin{array}{c|ccc} & x_1^{i_1}x_2^{i_2}x_3^{i_3} & \lambda_1^{j_1}\lambda_2^{j_2}\lambda_3^{j_3} & \mu_1^{k_1}\mu_2^{k_2}\mu_3^{k_3} \\ \hline [p]_2 & \alpha\beta\gamma & \alpha & \beta \\ [n+1]_2 & \alpha & 0 & 0 \\ [m+1]_2 & \beta & 0 & 0 \end{array} \tag{15}$$

The conditions on i_1, i_2, i_3 are given in (14); those on j_1, j_2, j_3 and k_1, k_2, k_3 are

$$j_1+j_2+j_3 = p-m-2;\ k_1+k_2+k_3 = p-n-2,\ k_1 < m. \tag{16}$$

We wish to show that, for the simplified forms (13), every row and column of D has one element 1 with all other elements 0. Thus $D = \pm 1$ and, in general, $D \neq 0$.

We prove this first for the rows. The first set of $[p]_2$ equations, obtained from the identity (I), arises from combinations $y_1^{l_1}y_2^{l_2}y_3^{l_3}$ with $l_1+l_2+l_3 = p-2$. These combinations divide into three aggregates:

(a) $y_1^{m-1-i_1}y_2^{n-1-i_2}y_3^{p-1-i_3}$ found in $J(x^{m+n-1}y^{p-2})$;

(b) $y_1^{j_1+m}y_2^{j_2}y_3^{j_3}$ found in $(\lambda y)^{p-2-m}\cdot y_1^m$;

(c) $y_1^{k_1}y_2^{k_2+n}y_3^{k_3}$ found in $(\mu y)^{p-2-n}\cdot y_2^n$.

A glance at the conditions in (14) and (16) enables us to draw at once the following conclusions. If $l_1 < m$ and $l_2 < n$, the combination is found once and only once in (a). If $l_1 < m$ and $l_2 \geqq n$, the combination is found once and only once in (c). If $l_1 \geqq m$ the combination is found once and only once in (b). Thus there is only one term in each of these $[p]_2$ equations. Evidently the other equations, obtained by multiplying either x_1^m or x_2^n by proper combinations of x, also contain only one term. Turning to the columns, it is clear that $x_1^{i_1}x_2^{i_2}x_3^{i_3}$ with the restrictions (16), will appear once and only once, with a coefficient α if $i_1 \geqq m$, with a coefficient β if $i_2 \geqq n$, whereas it will appear in J if both $i_1 < m$ and $i_2 < n$. Similarly, a coefficient λ, or a non-zero coefficient μ, will occur with one and only one combination of the y's so that the last two sets of columns have only one non-zero element.

With similar verifications for the other values of S we may assert that

(17) *The determinants D in* **3** *do not vanish identically.*

6. Determination of the resultant *R*. The identities which furnish the equations whose determinant is D all are expressed in terms of the given forms and of z. Thus D itself is a non-vanishing covariant. These equations are consistent if x is a common solution of the given forms whence D must contain the irreducible invariant R as a factor. The remaining covariant factor has a degree d in the coefficients of $(\alpha y)^m$ and an order md in the coordinates z. It must therefore be the d-th power of $(\alpha z)^m$. Hence

(18) *The determinants D listed in* **3** *are products of R and the powers of $(\alpha z)^m$ indicated.*

For sufficiently generic $(\alpha y)^m$, say

$$(\alpha y)^m = Ay_1^m + \cdots \qquad (A \neq 0) \tag{19}$$

and the choice of $z = 1, 0, 0 = E_1$ so that $(\alpha z)^m = A$, these determinants D can be materially simplified. We have employed this simplification in part in the particular case treated in **5**. More generally we use the identity (I) with the additional requirement that $(\mu y)^{k-n} = 0$ and $(\nu y)^{k-p} = 0$ have multiple points of orders $k - n - m + 1$ and $k - p - m + 1$ respectively at the reference point E_1. Thus the identities (IIa), (IIb) are eliminated by the fact that $(\mu y)^{k-n}$ and $(\nu y)^{k-p}$ have a smaller number of coefficients which must be eliminated. Similarly, we use the identities (6) with the additional requirement that $(\sigma y)^{h-n} = 0$ and $(\tau y)^{h-p} = 0$ have multiple points of orders $h - n - m + 1$ and $h - p - m + 1$ respectively at E_1. Then the equations

of the dialytic array obtained from the zero coefficients of σ and τ are dropped, and the remaining equations are linearly independent. Hence

(20) *If* $(\alpha y)^m = Ay_1{}^m + \cdots$ *with* $A \neq 0$, *and if the coefficients of the terms of degree greater than* $m-1$ *in* y_1 *of the forms* $(\mu y)^{k-n}$, $(\nu y)^{k-p}$, $(\sigma y)^{h-n}$, $(\tau y)^{h-p}$ *are all zero, then the elimination of the combinations* x^h, *and of* λ, μ, ν *from the equations furnished by the identity* (I), *and the additional dialytic equations furnished by* (6), *yields a determinant* D' *which is* $R \cdot A^d$ *where* d *is the power indicated in the table of Section* **3.**

It may be of interest to compare this result with more conventional methods. Let $m, n, p = 2, 3, 7$ and thus $l = 9$. The best results are obtained from median values of S. We use therefore the identity (I) for $J(x^4y^5)$, dropping the coefficient μ_1 of $y_1{}^2$ in $(\mu y)^2$. We obtain a determinant D' of order 30 which is $R \cdot A$. On the other hand an extreme value of S as in $J(x^0, y^9)$ yields a determinant of order 55 which is $R \cdot A^{16}$. These are to be compared with Van der Waerden's R as the G. C. D. of three determinants of order 66, or Macaulay's $R = D/D'$ where D, D' are determinants of order 66 and 25 respectively. Obviously the use of J improves very considerably both the order of D and the nature of the factor extraneous to R in D.

7. The unique apolar form. The three forms $(ay)^m$, $(\beta y)^n$, $(\gamma y)^p$, if sufficiently generic, and if in particular $R \neq 0$, have a unique apolar form of class l, $(a\eta)^l$, for which

$$(21) \qquad (\alpha a)^m (a\eta)^{l-m} \equiv 0, \quad (\beta a)^n (a\eta)^{l-n} \equiv 0, \quad (\gamma a)^p (a\eta)^{l-p} \equiv 0.$$

This apolar form can be obtained by bordering properly the determinants D of the table in **3**, or the determinants D' contemplated in (20). We first write the identity (I) involving $J(x^h, y^k)$ with an additional term $c(\eta y)^k$, and then add the equation, $(\xi x)^h = 0$. When $R = 0$, and there is a common point x, the new equations are all satisfied by arbitrary values η (when $c = 0$), provided only that ξ is on the common point x. Thus the determinant B or B', which arises from D or D' respectively by adding a column of η^k's and a row of ξ^h's to border the original determinant with respect to the matrix of its elements which arises from $J(x^h, y^k)$, contains a factor $(\xi x)^h$ when $R = 0$. If we take the reciprocal determinant [cf. (11)] formed for $J(x^k, y^h)$ and border it similarly, it contains a factor $(\xi x)^k$ when $R = 0$. But these reciprocal determinants are conjugate. Hence if either is bordered with both a column and a row of ξ's it will be a multiple of $(\xi x)^{h+k}$ when $R = 0$. But if x is a common point,

$(\xi x)^{h+k}$ is the unique apolar form. Since B or B' is of lower degree than the irreducible invariant R, it must also be the unique apolar form when $R \neq 0$. The presence of a factor $[(\alpha z)^m]^d$ in B or A^d in B' will not alter the argument so that we may state the theorem:

(22) *If the determinants D of the table in* **3**, *or the determinants D' of* (20), *are bordered with respect to the elements in the matrix which arises from $J(x^h, y^k)$ by a column of combinations η^k and a row of combinations η^h, the bordered determinants are expressions for the unique apolar form, $(a\eta)^l$, to within the same factor as appears in the original determinant extraneous to R.*

Thus the unique apolar form may be described as the *evectant* of R with respect to the coefficients of J in the determinant forms of R. Since the converse of the evectant process is the apolarity process we may assert that

(23) *The resultant R of the three given ternary forms is the apolarity invariant of the form $(a\eta)^l$ apolar to the given forms and of the jacobian $(\alpha\beta\gamma)(\alpha y)^{m-1}(\beta y)^{n-1}(\gamma y)^{p-1}$ of the given forms.*

This unique apolar form of degrees $np-1$, $mp-1$, $mn-1$ in the coefficients of the three given forms respectively is discussed much more completely in the following memoir.

The above account, taken in conjunction with (4), covers all cases of ternary elimination. We proceed to cover also all cases of quaternary elimination though no attempt will be made to give more than one determinant in a particular case.

8. Quaternary elimination with $D = R$. With four given quaternary forms, $(\alpha y)^m$, $(\beta y)^n$, $(\gamma y)^p$, $(\delta y)^q$ and

$$(24) \qquad m \leq n \leq p \leq q, \quad l = m + n + p + q - 4,$$

Morley and Coble [4] used the identity,

$$(\text{I}) \qquad J(x^h, y^k) + (\lambda_1 y)^{k-m} \cdot (\alpha y)^m + \cdots + (\lambda_4 y)^{k-q} \cdot (\delta y)^q \equiv 0 \qquad (h + k = l),$$

valid when x is a common point of the four corresponding surfaces, to obtain pure resultants $D = R$. However their exposition was confined to those cases in which the unknown coefficients $(\lambda_i y)$ in (I) were unique, i.e., when $k < m + n$. It was also confined to those cases in which additional equations

obtained from the dialytic array did not overlap, i. e., when $h < m + n$. Thus $h + k = l = m + n + p + q - 4 < 2m + 2n - 1$ or $p + q - n \leq m + 2$. In view of the inequalities (24) the cases thus covered are merely

$$(25)\qquad \begin{aligned} m, n, p, q &= m, m, m, m + r \ (r = 0, 1, 2);\\ &= m, m, m + 1, m + 1;\\ &= m, m + r, m + r, m + r \ (r = 1, 2);\\ &= m, m + 1, m + 1, m + 2. \end{aligned}$$

9. Quaternary elimination with $D' = R \cdot A^d$. Following the new ternary procedure developed above we now permit the overlapping of α with β, or γ, or δ, but not that of β with γ. This leads to the restrictions $h < n + p$ and $k < n + p$. These, in conjunction with (24), yield the following cases:

$$(26)\qquad m \leq n \leq p \leq q \leq n + p - m + 2.$$

We give only one determinant to cover these cases rather than the variety used in the ternary case. We use the identity

$$(\mathrm{I})\quad J(x^{m+q-3}, y^{n+p-1}) + (\lambda_1 y)^{n+p-m-1} \cdot (\alpha y)^m + \cdots + (\lambda_4 y)^{n+p-q-1} (\delta y)^q \equiv 0,$$

and form the determinant D' for $z = E_1 = 1, 0, 0, 0$. We shall reduce the number of coefficients in $(\lambda_2 y)^{p-1}$, $(\lambda_3 y)^{n-1}$, $(\lambda_4 y)^{n+p-q-1}$ by requiring the vanishing of all the coefficients which appear in the m-th polar of z. Then the identity (I) has unique coefficients λ. We also add the dialytic identities,

$$(\mathrm{III})\quad (\sigma_1 x)^{q-3} \cdot (\alpha x)^m \equiv 0 \ (\text{in } \sigma_1), \cdots, (\sigma_4 x)^{m-3} \cdot (\delta x)^q \equiv 0 \ (\text{in } \sigma_4),$$

with again the restriction that the coefficients of $(\sigma_2 y)^{m+q-n-3}$, $(\sigma_3 y)^{m+q-p-3}$ which appear in the m-th polar of z are themselves zero. Then the dialytic identities (III) yield independent equations. After these preliminaries the elimination of the combinations x of degree $m + q - 3$ and of the remaining coefficients λ yields a determinant D' of the form:

(27)

	a (x's)	a_1 (λ_1's)	a_2 (λ_2's)	a_3 (λ_3's)	a_4 (λ_4's)
(I) : b	$\alpha\beta\gamma\delta$	α	β	γ	δ
(IIIa) : b_1	α	0	0	0	0
(IIIb) : b_2	β	0	0	0	0
(IIIc) : b_3	γ	0	0	0	0
(IIId) : b_4	δ	0	0	0	0 .

The numbers, a, of columns and, b, of rows are as follows:

$$(28)\quad \begin{aligned} a &= [m+q]_3, & b &= [n+p+2]_3,\\ a_1 &= [n+p-m+2]_3, & b_1 &= [q]_3,\\ a_2 &= [p+2]_3-[p-m+2]_3, & b_2 &= [m+q-n]_3-[q-n]_3,\\ a_3 &= [n+2]_3-[n-m+2]_3, & b_3 &= [m+q-p]_3-[q-p]_3,\\ a_4 &= [n+p-q+2]_3 & b_4 &= [m]_3.\\ &\quad -[n+p-q-m+2]_3, & & \end{aligned}$$

It can now be verified that

$$a+a_1+\cdots+a_4=b+b_1+\cdots+b_4.$$

The degree in the coefficients of $J(x^h, y^k)$ is

$$j=b-a_1-\cdots-a_4=a-b_1-\cdots-b_4.$$

The degrees of the expansion of D' in the coefficients $\alpha, \beta, \gamma, \delta$ are respectively $npq+d$, mpq, mnq, mnp, where

$$(29)\quad d=[p-m+2]_3+[n-m+2]_3+[n+p-q-m+2]_3 \\ +[q-n]_3+[q-p]_3.$$

Then the usual verification that $D'\not\equiv 0$ yields the theorem:

(30) *If* $(\alpha y)^m = Ay_1^m+\cdots$ *with* $A\neq 0$, *the determinants* D' *indicated in* (27) *is* $R\cdot A^d$ *with* d *as in* (29).

Naturally if z had been chosen generically and the determinant D' had been replaced by a determinant D formed as in the ternary case, the value of D would be $R\cdot[(\alpha z)^m]^d$.

10. Quaternary elimination in the general case. We are here concerned with orders

$$(31)\quad m\leq n\leq p\leq q\geq n+p-m+3,$$

these being the cases not covered in **8** and **9**. In order to keep the dialytic equations as simple as possible we shall take $h=m+n-1$. The identity (I) then reads:

$$(\mathrm{I}):\quad J(x^{m+n-1}, y^{p+q-3})+(\lambda_1 y)^{p+q-m-3}\cdot(\alpha y)^m+\cdots+(\lambda_4 y)^{p-3}\cdot(\delta y)^q\equiv 0.$$

Since $\lambda_2, \lambda_3, \lambda_4$ can be affected by multiples of $(\alpha y)^m$, we restore their uniqueness in this respect by the requirement that

$$(\mathrm{II}):\quad (\lambda_2 y)^{p+q-n-m-3}(\lambda_2 z)^m\equiv 0,\ (\lambda_3 y)^{q-m-3}(\lambda_3 z)^m\equiv 0, \\ (\lambda_4 y)^{p-m-3}(\lambda_4 z)^m\equiv 0.$$

Since λ_3, λ_4 can also be affected by multiples of $(\beta y)^n$ we restore their uniqueness in this respect by the requirement that

$$\text{(III)}: \quad \begin{aligned} &(\lambda_3 y)^{q-n-3}(\lambda_3 z')^n + (\lambda_{23} y)^{q-m-n-3} \cdot (\alpha y)^m \equiv 0, \\ &(\lambda_4 y)^{p-n-3}(\lambda_4 z')^n + (\lambda_{24} y)^{p-m-n-3} \cdot (\alpha y)^m \equiv 0. \end{aligned}$$

Finally we add the dialytic equations,

$$\text{(IV)}: \quad (\eta x)^{n-1} \cdot (\alpha x)^m \equiv 0, \quad (\eta x)^{m-1} \cdot (\beta x)^n \equiv 0,$$

on the assumption that $p \geqq m + n$. This assumption is by no means necessary but it is more general than $p < m + n$ and it leads at least to the exclusion of $(\gamma x)^p$ from the dialytic equations and thus yields a more specific case.

The general effect of restricting h to $m + n - 1$ is to throw the inherent difficulty of the problem entirely upon the form of the conditions which make the coefficients $\lambda_1, \cdots, \lambda_4$ in (I) unique. We are using here the following property of a module of j forms in j variables with $R \neq 0$: If $z, z', z'', \cdots$ are arbitrarily chosen value systems of y, the form,

$$(\lambda_1 y)^{k-l_1} \cdot (\alpha_1 y)^{l_1} + (\lambda_2 y)^{k-l_2} \cdot (\alpha_2 y)^{l_2} + \cdots + (\lambda_j y)^{k-l_j} \cdot (\alpha_j y)^{l_j},$$

belonging to the module has unique coefficients $\lambda_1, \lambda_2, \cdots, \lambda_j$ if these coefficients are subject to the conditions that the l_1-th polars of z as to $\alpha_2, \cdots, \alpha_j$ vanish identically, the l_2-th polars of z' as to $\alpha_3, \cdots, \alpha_j$ are in the module determined by $(\alpha y)^{l_1}$, the l_3-th polars of z'' as to $\alpha_4, \cdots, \alpha_j$ are in the module determined by $(\alpha_1 y)^{l_1}$, $(\alpha_2 y)^{l_2}$, etc. We mention this general property here partly because we are using a very special case of it and partly because the following memoir contains a radically different method for obtaining unique coefficients λ.

With the equations derived from the identities (I), (II), (III), (IV) we eliminate the combinations x^h, and the coefficients $\lambda_1, \cdots, \lambda_4, \lambda_{23}, \lambda_{24}$ to obtain the determinant D,

(32)

	$[m+n+2]_3$ x	$[q+p-m]_3$ λ_1	$[q+p-n]_3$ λ_2	$[q]_3$ λ_3	$[p]_3$ λ_4	$[q-m-n]_3$ λ_{23}	$[p-m-$ λ_{24}
$[q+p]_3$	$\alpha\beta\gamma\delta$	α	β	γ	δ	0	0
$[q+p-m-n]_3$	0	0	z^m	0	0	0	0
$[q-m]_3$	0	0	0	z^m	0	0	0
$[p-m]_3$	0	0	0	0	z^m	0	0
$[q-n]_3$	0	0	0	$z^{n\prime}$	0	α	0
$[p-n]_3$	0	0	0	0	z'^n	0	α
$[n+2]_3$	α	0	0	0	0	0	0
$[m+2]_3$	β	0	0	0	0	0	0

The degrees in the coefficients of $(\gamma y)^p$, $(\delta y)^q$ of this determinant are mnq and mnp respectively. The degrees in the coefficients of $(\alpha y)^m$, $(\beta y)^n$ are respectively $npq + d$ and $mpq + e$ where

$$\begin{aligned} &d = [q+p-m-n]_3 + [q-m]_3 + [p-m]_3 \\ &e = [q-n]_3 - [q-n-m]_3 + [p-n]_3 - [p-n-m]_3. \end{aligned} \tag{33}$$

Thus we might expect D to be $R \cdot \{[(\alpha z)^m]^d \cdot [(\beta z')^n]^e + \cdots\}$. We shall actually determine the nature of this factor, extraneous to R, in D and thus prove incidentally that $D \neq 0$ and that its significant factor is R.

Since we may, and will henceforth, take z, z' to be

$$z = 1, 0, 0, 0; \quad z' = 0, 1, 0, 0$$

the determination of the extraneous factor E is essentially a binary problem which can be accomplished in the particular case,

$$\begin{aligned} &(\alpha y)^m = A_0 y_1^m + [m]_1 A_1 y_1^{m-1} y_2 + \cdots + A_m y_2^m \quad (A_0 \neq 0), \\ &(\beta y)^n = B_0 y_1^n + [n]_1 B_1 y_1^{n-1} y_2 + \cdots + B_n y_2^n, \\ &(\gamma y)^p = y_3^p, \\ &(\delta y)^q = y_4^q. \end{aligned}$$

If, for this particular case, the value of E is determined in terms of A_i, B_j its value for D in (32) will be obtained by setting

$$A_i = (\alpha z)^{m-i} (\alpha z')^i, \qquad B_j = (\beta z)^{n-j} (\beta z')^j. \tag{34}$$

In general we arrange terms $y_1^{i_1} y_2^{i_2} y_3^{i_3} y_4^{i_4}$ (or corresponding coefficients or equations) first with reference to descending i_1, and, for equal i_1's, with reference to descending i_2, etc. We refer to this briefly as a *canonical arrangement*.

Let us first trim off outlying matrices from D. The simplest ones to dispose of are those in z^m. With $z = 1, 0, 0, 0$, the identities (II) merely assert that all terms in λ_2, λ_3, λ_4 with $i_1 \geq m$ have zero coefficients so that our first reduction can be described as

$$R_1: \ \lambda_2, \lambda_3, \lambda_4 \text{ have only terms for which } i_1 < m.$$

We consider now the matrices in z'^n and α under λ_4 and λ_{24} respectively derived from the $[p-n]_3$ equations implied by the second identity in (III). These equations divide into two classes: (a) those for which $i_1 \geq m$, and (b) those for which $i_1 < m$. In the equations (a) the coefficients λ_4 are zero due to R_1 and a particular equation must contain a term of $(\lambda_{24} y)^{p-n-m-3} \cdot A_0 y_1^m$.

If then the equations (a) are arranged canonically, they will have under λ_{24}, also in canonical arrangement, a determinant with a diagonal made up of elements A_0 and with zeros to the right of the diagonal. This determinant, $A_0^{[p-n-m]_3}$, is a factor of D. With the matrix α on the right thus eliminated, the remaining equations (b) merely require the vanishing of those coefficients of λ_4 for which $i_2 \geqq n$. The same considerations apply to the matrices in z'^n and α under λ_3 and λ_{23} respectively so that, for both sets of matrices, we acquire an outstanding factor of D,

$$E_1 = A_0^{[p-n-m]_3+[q-n-m]_3},$$

and have a reduction:

$$R_2: \begin{array}{l} \lambda_2 \text{ has only terms for which } i_1 < m; \\ \lambda_3, \lambda_4 \text{ have only terms for which } i_1 < m \text{ and } i_2 < n. \end{array}$$

The determinant which results from these reductions has the form,

$$\begin{array}{r|ccccc} & [m+n+2]_3 & [q+p-m]_3 & a_2 & a_3 & a_4 \\ & x & \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \\ \hline [q+p]_3 & \alpha\beta\gamma\delta & \alpha & \beta & \gamma & \delta \\ D' = [n+2]_3 & \alpha & 0 & 0 & 0 & 0 \\ [m+2]_3 & \beta & 0 & 0 & 0 & 0, \end{array}$$

where

$$a_2 = [q+p-n]_3 - [q+p-n-m]_3,$$
$$a_3 = [q]_3 - [q-m]_3 - [q-n]_3 + [q-m-n]_3 = mn[q-1-(m+n)/2]$$
$$a_4 = [p]_3 - [p-m]_3 - [p-n]_3 + [p-m-n]_3 = mn[p-1-(m+n)/2]$$

The $[q+p]_3$ equations derived from the identity (I) may be characterized by i_1, i_2, i_3, i_4, (the coefficient of $y_1^{i_1}y_2^{i_2}y_3^{i_3}y_4^{i_4}$ in (I)) where $i_1 + i_2 + i_3 + i_4 = q + p - 3$. We isolate from these two sets of equations characterized as follows:

$$\text{Eq. (A)}: i_3 \geqq p,\ i_1 \leqq m-1,\ i_2 \leqq n-1;$$
$$\text{Eq. (B)}: i_4 \geqq q,\ i_1 \leqq m-1,\ i_2 \leqq n-1.$$

The two sets do not collide because $i_3 + i_4 \leqq p + q - 3$. In the identity (I) for our special case, the terms in $\lambda_3^{q-3} \cdot y_3^p + \lambda_4^{p-3} \cdot y_4^q$ occur in one and only one of these equations, and each of the equations contains one and only one coefficient of either λ_3 or λ_4. No other equations contain coefficients of λ_3 or λ_4. Thus the matrices under λ_3, λ_4 contribute numerical factors of D' which

may be dropped if λ_3, λ_4 and Eq. (A), (B) are dropped. Then D' reduces to D'' where

$$D'' = \begin{array}{r|ccc} & [m+n+2]_3 & [q+p-m]_3 & a_2 \\ & x & \lambda_1 & \lambda_2 \\ \hline [q+p]_3 - a_3 - a_4 & \alpha\beta\gamma\delta & \alpha & \beta \\ [n+2]_3 & \alpha & 0 & 0 \\ [m+2]_3 & \beta & 0 & 0 \end{array} = D/E_1.$$

We proceed now to account for the matrices in the first columns of D'' by isolating a set of equations (C) from the identity characterized as follows:

$$\text{Eq. (C)}: \; i_1 + i_2 \leq m + n - 2, \; i_3 \leq p - 1, \; i_4 \leq q - 1.$$

These are the only equations which involve the coefficients of $J(x^{m+n-1}, y^{p+q-3})$. For $J(y^{m+n+p+q-4})$ has, in our special case, the value,

$$J(y) = \begin{vmatrix} (\alpha y)^{m-1}\alpha_1 & (\alpha y)^{m-1}\alpha_2 \\ (\beta y)^{n-1}\beta_1 & (\beta y)^{n-1}\beta_2 \end{vmatrix} y_3^{p-1} y_4^{q-1}.$$

Since $(\alpha y)^m$, $(\beta y)^n$ are binary forms in y_1, y_2 alone, the highest degree in y_1, y_2 which can be attained after polarization as to x is $m + n - 2$, and this and all lower degrees in y_1, y_2 occur in $J(x, y)$. We note also that the combinations of x_1, x_2 to the degree $m + n - 1$ will not occur in $J(x, y)$. These $m + n$ combinations arise from the matrices α, β below J by using the identities (IV) for $\xi_3 = \xi_4 = 0$. This particular set of $m + n$ equations in these $m + n$ combinations has for determinant Sylvester's dialytic determinant, the resultant

$$r = r(A^m, B^n)$$

of the binary forms, $A_0 y_1^m + \cdots, B_0 y_1^n + \cdots$. Thus r^1 is a factor of D''.

In order to handle the remaining dialytic equations and the equations C we recall the $m + n - 1$ various forms of the binary eliminant of $(\alpha y)^m$, $(\beta y)^n$ which involve the coefficients of their jacobian [cf. [5]]. These can be tabulated in the following typical sequence:

$$1^\circ: \; J(x^{m+n-2}) = 0, \qquad \qquad \begin{array}{r|c} & x^{m+n-2} \\ \hline 1 & \alpha\beta \\ n-1 & \alpha \\ m-1 & \beta \end{array} = r;$$
$$(\xi x)^{n-2} \cdot (\alpha x)^m \equiv 0, \; (\xi x)^{m-2} \cdot (\beta x)^n \equiv 0,$$

. . . .

2

$$(m-1)^0:\ J(x^n, y^{m-2}) \equiv 0, \qquad (\xi x)^{n-m}\cdot(\alpha x)^m \equiv 0,\ (\beta x)^n = 0, \qquad \begin{array}{r|l} & x^n \\ \hline m-1 & \alpha\beta \\ n-m+1 & \alpha \\ 1 & \beta \end{array} = r;$$

$$m^0:\ J(x^{n-1}y^{m-1}) \equiv 0, \qquad (\xi x)^{n-m-1}\cdot(\alpha x)^m \equiv 0, \qquad \begin{array}{r|l} & x^{n-1} \\ \hline m & \alpha\beta \\ n-m & \alpha \end{array} = r;$$

$$(m+1)^0:\ J(x^{n-2}, y^m) + \lambda(\alpha y)^m \equiv 0, \qquad (\xi x)^{n-m-2}\cdot(\alpha x)^m \equiv 0, \qquad \begin{array}{r|ll} & x^{n-2} & \lambda^0 \\ \hline m+1 & \alpha\beta & \alpha \\ n-m-1 & \alpha & 0 \end{array} = r;$$

. . . .

$$(n-1)^0:\ J(x^m, y^{n-2}) + (\lambda y)^{n-m-2}\cdot(\alpha y^m \equiv 0, \qquad (\alpha x)^m = 0, \qquad \begin{array}{r|ll} & x^m & \lambda^{n-m-2} \\ \hline n-1 & \alpha\beta & \alpha \\ 1 & \alpha & 0 \end{array} = r;$$

$$n^0:\ J(x^{m-1}, y^{n-1}) + (\lambda y)^{n-m-1}\cdot(\alpha y)^m \equiv 0, \qquad \begin{array}{r|ll} & x^{m-1} & \lambda^{n-m-1} \\ \hline n & \alpha\beta & \alpha \end{array} = r;$$

$$(n+1)^0:\ J(x^{m-2}, y^n) + (\lambda y)^{n-m}\cdot(\alpha y)^m + \mu(\beta y)^n \equiv 0,$$

$$\begin{array}{r|lll} & x^{m-2} & \lambda^{n-m} & \mu^0 \\ \hline n+1 & \alpha\beta & \alpha & \beta \end{array} = r;$$

. . . .

$$(n+m-1)^0:\ J(y^{m+n-2}) + (\lambda y)^{n-2}\cdot(\alpha y)^m + (\mu y)^{m-2}\cdot(\beta y)^n \equiv 0,$$

$$\begin{array}{r|lll} & 1 & \lambda^{n-2} & \mu^{m-2} \\ \hline m+n-1 & \alpha\beta & \alpha & \beta \end{array} = r.$$

The binary identity 1^0 contains all combinations of x_1, x_2 to the degree $m+n-2$. To bring these up to the degree $m+n-1$ of J in x we can multiply them by x_3 or by x_4. If x_3 is the factor chosen for all of the equations 1^0, then those arising from J must have a unique further factor $y_3^{p-2}y_4^{q-1}$ in order to produce equations (C); if x_4 is chosen, the further

factor is $y_3^{p-1}y_4^{q-2}$. Corresponding to these two factors we have two sets of equations 1^c which have determinants r and the case 1^0 contributes a factor r^2 of D''. In the next identity 2^0 we must use factors $x_3^2y_3^{p-3}y_4^{q-1}$, $x_3x_4y_3^{p-2}y_4^{q-2}$, $x_4^2y_3^{p-1}y_4^{q-3}$, and thus obtain a factor r^3. Proceeding in this way up to $(m+n-1)^0$ we obtain a factor

$$E_2 = r^{1+2+3+\ldots+(m+n)} = r^{[n+m+1]_2}$$

of D.

In the above process we have used all of the equations (C) and all of the dialytic equations obtained from the identities (IV) and have thus exhausted the matrices in the first column of D. Of the variables we have eliminated all of the x^{m+n-1}. In addition, however, from $(m+1)^0$ on we have eliminated coefficients λ_1 (as binary coefficients λ) and coefficients λ_2 (as binary coefficients μ). The coefficients of $(\lambda_1 y)^{q+p-m-3}$ thus eliminated are characterized by

$$(35) \qquad i_1+i_2 \leq n-2,\ i_3 \leq p-1,\ i_4 \leq q-1,$$

and they number

$$\begin{aligned} b_1 &= (m+2)\cdot 1 + (m+3)\cdot 2 + \cdots + (m+n)\cdot(n-1) \\ &= m[n]_2 + 2[n+1]_3; \end{aligned}$$

those of $(\lambda_2 y)^{q+p-n-3}$ eliminated are characterized by

$$(36) \qquad i_1+i_2 \leq m-2,\ i_3 \leq p-1,\ i_4 \leq q-1,$$

and they number

$$\begin{aligned} b_2 &= (n+2)\cdot 1 + (n+3)\cdot 2 + \cdots + (n+m)(m-1) \\ &= n[m]_2 + 2[m+1]_3. \end{aligned}$$

The remaining equations and variables have a determinant of the form,

$$D''' = c \left| \begin{array}{cc} c_1(\lambda_1\text{'s}) & c_2(\lambda_2\text{'s}) \\ \hline \alpha & \beta \end{array} \right| = D/E_1E_2,$$

where

$$\begin{aligned} c_1 &= [q+p-m]_3 - b_1, \\ c_2 &= a_2 - b_2, \\ c &= [q+p]_3 - a_3 - a_4 - b_1 - b_2 - [m+n+2]_3 \\ &\qquad + [n+2]_3 + [m+2]_3. \end{aligned}$$

The equations are those which arise from the identity (I) except (A), (B), (C). However, for these the identity (I) now takes the simple form

$$(\mathrm{I}'): \quad (\lambda_1 y)^{q+p-m-3}\cdot(\alpha y)^m + (\lambda_2 y)^{q+p-n-3}\cdot(\beta y)^n \equiv 0.$$

In this all of the coefficients of λ_1 and λ_2 are present except those excluded by (C) as described in (35), (36) and also except that λ_2 contains terms in $y_1^0, y_1^1, \cdots, y_1^{m-1}$ only. Consider then a new set of equations (D) derived from (I′) and characterized by

$$\text{Eq. (D)}: \ i_1 \geqq m+n.$$

These contain only the coefficients λ_1 in (I′). If then both the equations and the coefficients λ_1 which they contain are arranged in the canonical order their determinant will have elements A_0 on the diagonal and zeros to the right of the diagonal. The number of these equations is

$$[2]_2 + [3]_2 + \cdots + [q+p-1-m-n]_2 = [q+p-m-n]_3.$$

Thus D''' contains a factor,

$$E_3 = A_0^{[q+p-m-n]_3}.$$

The remaining equations from I′ are those, exclusive of (A), (B), (C) which would be obtained from an identity of the form,

(I″):

$$\begin{aligned}&[(\lambda_{1,0}y)^{q+p-m-n-2}y_1^{n-1} + \cdots + (\lambda_{1,n-1}y)^{q+p-m-3}][A_0y_1^m + \cdots + A_my_2^m]\\ &+ [(\lambda_{2,0})y^{q+p-m-n-2}y_1^{m-1} + \cdots + (\lambda_{2,m-1}y)^{q+p-n-3}][B_0y_1^n + \cdots + B_ny_2^n] \equiv 0,\end{aligned}$$

where the $\lambda_{i,j}$ are forms in y_2, y_3, y_4 with coefficients excluded by (35), (36) omitted.

In the permissible equations derived from (I″) let i_3, i_4 be fixed. Then, in the forms $\lambda_{i,j}$, the exponent of y_2 is also fixed. Thus we have, for $i_1 + i_2 = q+p-3-i_3-i_4$ and $i_1 = 0, \cdots, m+n-1$, precisely $m+n$ equations to fix the values of the $m+n$ coefficients $\lambda_{i,j}(i_1, i_2, i_3, i_4)$. These $m+n$ coefficients will appear in no other equations and their determinant is the resultant, r. This factor will appear for all sets of $m+n$ equations from $i_1 + i_2 = m+n-1$ up to $i_1 + i_2 = q+p-3$ with corresponding $i_3 + i_4$ descending from $q+p-m-n-2$ down to 0. Thus the number of such sets of $m+n$ equations (i. e., the number of sets i_3, i_4) is

$$1 + 2 + \cdots + (q+p-m-n-1) = [q+p-m-n]_2.$$

We have therefore another factor of D''', namely

$$E_4 = r^{[q+p-m-n]_2}.$$

Consider now the equations from (I″) for fixed i_3, i_4 when $i_1 + i_2 = m+n-2$. We lose two variables, one from $(\lambda_{1,0}y)$ and one from $(\lambda_{2,0}y)$.

But we also lose two equations, one because the range of i_1 from 0 to $m+n-2$ is reduced, and one from the combination $i_1, i_2 = m-1, n-1$. Indeed this latter equation is barred by (A) when $i_3 \geqq p$, by (B) when $i_4 \geqq q$, and by (C) for intermediate values. Thus we have, for fixed i_3, i_4 precisely $m+n-2$ equations in $m+n-2$ coefficients whose determinant is

$$r_1 = r_1(A^{n-1}, B^{m-1}),$$

a *first sub-resultant of* r, obtained by dropping two rows and two columns of r in the dialytic form of r. The number of pairs i_3, i_4 which contribute such a factor r_1 of D''' is the number not excluded by (C), i. e., $p+q-2(m+n)$.

When $i_1 + i_2 = m+n-3$, we lose two equations because of the range for i_1, and two equations for $i_1, i_2 = m-1, n-2$; $m-2, n-1$, and thus have, for fixed $i_3, i_4, m+n-4$ equations in $m+n-4$ coefficients whose determinant is

$$r_2 = r_2(A^{n-2}, B^{m-2}),$$

a *second sub-resultant of* r. Since the value of $i_3 + i_4$ is now increased by one we have $p+q-2(m+n-1)$ admissible pairs i_3, i_4. This process goes on until $i_1 + i_2 = n$, and we get a subresultant $r_{m-1}(A^{n-m+1}, B^1)$ for each of $p+q-2(n+2)$ pairs i_3, i_4. Thus we have a further factor of D''' namely,

$$E_5 = r_1{}^{p+q-2(m+n)} r_2{}^{p+q-2(m+n-1)} \cdots r_{m-1}{}^{p+q-2(n+2)}.$$

When $i_1 + i_2 = n-1$, r_m is merely $A_0{}^{n-m}$ taken $p+q-2(n+1)$ times. Thereafter the number of equations lost is due only to the decreasing range of i_1. Finally, with $i_1 + i_2 = m$, we reach $A_0{}^1$ taken $p+q-2m-4$ times. When $i_1 + i_2 \leqq m-1$ the equations are all excluded by (A), (B), or (C). Thus we have a further final factor of D''', namely,

$$E_6 = A_0{}^f,$$

where

$$\begin{aligned} f &= (n-m)(p+q-2n-2) + (n-m-1)(p+q-2n) \\ &\qquad + \cdots + (1)(p+q-2m-4) \\ &= (p+q-2m-2)[n-m+1]_2 \\ &\qquad - (n-m)(n-m+1)(2n-2m+1)/3. \end{aligned}$$

We have thus evaluated the deterimnant D for the special case here used in the form,

$$D = E_1 E_2 E_3 E_4 E_5 E_6.$$

In this special case, however, R, the resultant of the four forms, is r^{pq}. This factor r appears in E_2 and E_4 so that we have finally

(37) $$D = R \cdot A_0{}^g r^h E_5,$$
where

$$g = f + [q - n - m]_3 + [p - n - m]_3 + [q + p - n - m]_3,$$
$$h = [q + p - m - n]_2 + [n + m + 1]_2 - pq.$$

We may state the above final result as follows:

(38) *The determinant D in (32) is the resultant R of the four given forms multiplied by an extraneous factor which is a product of powers of $(\alpha z)^m$, and of powers of the resultant and subresultants of two binary forms of orders m, n with coefficients* (34).

As a check on the entire procedure we had already found in (33) the degrees d, e of the extraneous factor in the coefficients of $(\alpha y)^m$, $(\beta y)^n$. Now the degree of the above extraneous factor $A_0{}^g r^h E_5$ in the coefficients of $(\alpha y)^m$ is

$$\begin{aligned} g + nh + \{ & (n - m + 1)(p + q - 2n - 4) \\ & + (n - m + 2)(p + q - 2n - 6) + \cdots \\ & + (n - 1)(p + q - 2n - 2m)\} \\ = g + & nh + (p + q - 2n - 2)\{(m - 1)(n - m) + [m]_2\} \\ & - m(m - 1)(2m - 1)/3 - 2(n - m)[m]_2, \end{aligned}$$

and this is equal to d in (33). Similarly the degree of $r^h E_5$ in the coefficients of $(\beta y)^n$ turns out to be e in (33).

11. Elimination in higher domains. In the Morley-Coble account [4], the restriction (a), $h \leq m + n - 1$, $k \leq m + n - 1$, barred all identities of the sort, $(\beta y)^n \cdot (\alpha y)^m \equiv (\alpha y)^m \cdot (\beta y)^n$, and yielded a pure resultant R in the cases covered. These cases comprised arbitrary orders m, n in the ternary domain, and arbitrary m in the quaternary domain. In the present account $(1, \cdots, 9)$ the restrictions (b), $h \leq n + p - 1$, $k \leq n + p - 1$ admit identities of the above sort only between $(\alpha y)^m$ and any one of the remaining forms. The cases then covered comprised arbitrary m, n, p in both the ternary and quaternary domains. Finally in **10** the restrictions (c), $h \leq p + q - 1$, $k \leq p + q - 1$ admit identites of the above sort between $(\alpha y)^m$ or $(\beta y)^n$ and any one of the remaining forms. The cases then covered exhausted the quarternary domain.

Nevertheless the flexibility of the method rapidly diminishes. The essential reason for this is that, in j variables, we have

$$h + k = l = l_1 + \cdots + l_j - j,$$

and, for given l, there is only one degree of freedom in the choice of h, k whereas there are $j-1$ degrees in the choice of the orders $l_1, \cdots, l_j$. Thus, in the quinary domain with $l_1, \cdots, l_5 = 4, 5, 5, 5, 5$; $l = 19$, the most favorable choice of h, k is 9, 10 and then all of the possible overlapping identities appear. This situation would require the introduction of four parametric points z, z', z'', z''' as explained in **10**. If, as further parameters are necessary, some well defined pattern for the extraneous factor would appear, it might be worth while to pursue the method more generally. Our main reason for the determination of E in **10** was not to complete definitely the ternary and quaternary elimination but rather to show that such a pattern is not to be expected.

UNIVERSITY OF ILLINOIS.

REFERENCES.

[1] B. L. Van der Waerden, *Moderne Algebra*, Part II, Berlin (Springer), 1931.

[2] F. S. Macaulay, *The Algebraic Theory of Modular Systems*, Cambridge Tracts, No. 19, Cambridge University Press, 1916.

[3] F. Morley, "The eliminant of a net of curves," *Americal Journal of Mathematics*, vol. 47 (1925), pp. 91-97.

[4] F. Morley and A. B. Coble, "New results in elimination," *American Journal of Mathematics*, vol. 49 (1927), pp. 463-488.

[5] T. W. Moore, "On the resultant of two binary forms," *Annals of Mathematics*, vol. 31 (1930), pp. 185-189.

ON THE EXPRESSION OF AN ALGEBRAIC FORM IN TERMS OF A SET OF FORMS WITH NON-ZERO RESULTANT.*

By Arthur B. Coble

1. Introduction. Given n algebraic forms in n variables,

$$f_i = (\alpha_i y)^{l_i} \quad (i = 1, \cdots, n), \quad (l_1 \leqq 2l_2 \leqq \cdots \leqq l_n),$$

we set

$$l = \Sigma_i (l_i - 1), \qquad L_i = (\prod_{j=1}^{j=n} l_j)/l_i.$$

If the resultant R of the n given forms does not vanish, it is well known that an algebraic form of order $m > l$ can be expressed as a member of the module determined by the given forms, i. e., that

$$R \cdot (\delta y)^m \equiv \sum_{i=1}^{i=n} (\beta_i y)^{m-l_i} \cdot (\alpha_i y)^{l_i}. \tag{1}$$

Ordinarily (1) is indicated by the notation $R \cdot (\delta y)^m \equiv 0$ (mod. $f_1, \cdots, f_n$), with no attention to the coefficients, $(\beta_i y)^{m-l_i}$, in the expression (1). It is the purpose of this article to determine explicit values for these coefficients as covariants of the given forms and of $(\delta y)^m$. We also determine an explicit extension of the module defined by the given forms which will yield an expression of type (1) for orders $0 < m \leqq l$. In these expressions the resultant R on the left of (1) will be balanced on the right by the coefficients of the unique form, $A = (a\eta)^l$, apolar to the n forms f_i which define the module. This balance is accentuated by the fact that, in the generic case, *neither R nor A* is explicitly known. Thus an independent discussion of the form A and of some of its properties may be in order. Again we find it convenient to denote the binomial coefficients by the symbol

$$[n]_k = {}_nC_k.$$

2. The unique form A apolar to the given forms. Let $C_{l_1}(r)$ be the number of linearly independent conditions imposed on a form $F = (b\eta)^r$ of class r by its apolarity with a given form, $f_1 = (\alpha_1 y)^{l_1}$, of order l_1, i. e., by the identical vanishing of $(b\alpha_1)^{l_1}(b\eta)^{r-l_1}$. Let $C_{l_1,\ldots,l_k}(r)$ be the number of linearly independent conditions imposed on F by its apolarity with k given generic forms $f_1, f_2, \cdots, f_k$. We wish to prove that

* Received May 18, 1946.

$$(2) \qquad C_{l_1,\ldots,l_k}(r) = \Sigma C_{l_1}(r) - \Sigma C_{l_1+l_2}(r) + \Sigma C_{l_1+l_2+l_3}(r) - \cdots + (-1)^{k-1} C_{l_1+\ldots+l_k}(r).$$

where the summations refer to all of the combinations indicated by the subscripts which can be drawn from $l_1, \cdots, l_k$.

We observe that the value (2) of $C_{l_1,\ldots,l_k}(r)$ can be separated into three parts as follows,

$$(3) \qquad C_{l_1,\ldots,l_k}(r) = C_{l_1,\ldots,l_{k-1}}(r) + C_{l_k}(r) - C_{l_1,\ldots,l_{k-1}}(r - l_k),$$

according as a subscript on the right of (2) does not contain l_k, contains l_k only, or contains l_k together with other l's. This formula (3) is a recursion formula for the determination of $C_{l_1,\ldots,l_k}(r)$ in terms of $C_{l_1,\ldots l_{k-1}}$. It expresses the fact that, of the C_{l_k} conditions imposed on F by its apolarity with f_k, conditions in themselves linearly independent, those which arise from the apolarity of $(b\alpha_k)^{l_k}(b\eta)^{r-l_k}$ with the $k-1$ earlier forms are themselves a consequence of the earlier apolarities. This recursion formula is then the basis for an induction proof of (2) which by definition is valid when $k = 1$.

We now make use of an identity among the binomial coefficients, namely:

$$(4) \qquad [l_1 + \cdots + l_n - 1]_{n-1} - \Sigma[l_1 + \cdots + l_{n-1} - 1]_{n-1} + \Sigma[l_1 + \cdots + l_{n-2} - 1]_{n-1} - \cdots (-1)^{n-1}\Sigma[l_1 - 1]_{n-1} + (-1)^n[-1]_{n-1} = 0.$$

All of the terms of this identity cancel due to relations of the form $(1-1)^j = 0$ $(j = 1, \cdots, n)$. The first term of (4) is the number of coefficients of the form $A = (a\eta)^l$ and the last term is -1. The intermediate terms by comparison with (2) are $-C_{l_1,\ldots,l_n}(l)$. Hence

(5) *If $f_i = (\alpha_i y)^{l_i}$ $(i = 1, \cdots, n)$ are n generic forms of the orders indicated in variables $y_1, \cdots, y_n$, there is a single form $A = (a\eta)^l$ of class $l = l_1 + \cdots + l_n - n$ which is apolar to the n given forms.*

That the conditions, which have not been barred above as obviously dependent on earlier conditions, are themselves linearly independent is evident from the very special, but sufficiently generic, case:

$$(6) \qquad f_i = y_i^{l_i} \quad (i = 1, \cdots, n).$$

For, then, $A = c\eta_1^{l_1-1}\eta_2^{l_2-1} \cdots \eta_n^{l_n-1}$, and the uniqueness is evident.

In line with similar definitions for the resultant of the forms f_i we may state that

(7) *For given forms, $(\alpha_i y)^{l_i}$ $(i=1,\cdots,n)$, the unique apolar form $A=(a\eta)^l$ is the* G. C. D. *of the determinants of order $[l-1]_{n-1}$ formed from the array of linear equations in the coefficients a obtained from the apolarity conditions, $(\alpha_i a)^{l_i}(a\eta)^{l-l_i}\equiv 0$, and the additional linear equation, $(a\eta)^l=0$.*

Indeed, all of the determinants formed from the identities alone must vanish since the coefficients of $(a\eta)^l$ are not all zero. Thus the non-zero determinants must involve the equation $(a\eta)^l=0$. A particular non-vanishing one will therefore be the apolar form multiplied by some factor which contains the coefficients of the given forms and which depends upon the manner of selecting equations from the identities.

Another definition of the apolar form is that A is some one of the non-zero determinants mentioned divided by the G. C. D. of the coefficients of the terms in η.

We now prove the theorem:

(8) *If $L_i=l_1l_2\cdots l_n/l_i$, the degree of the unique apolar form $(a\eta)^l$ in the coefficients of f_i is L_i-1.*

We prove this for the case $i=n$, making use of the identity,

$$(9)\qquad \begin{aligned}&[l_1+\cdots+l_{n-1}+m]_{n-1}-\Sigma[l_1+\cdots+l_{n-2}+m]_{n-1}\\ &\qquad+\Sigma[l_1+\cdots+l_{n-3}+m]_{n-1}-\cdots+(-1)^{n-1}[m]_{n-1}\\ &=L_n \qquad\qquad\qquad\qquad (m=l_n-1).\end{aligned}$$

The first term of this identity is the number of linearly independent forms of class l. The remaining terms on the left are $-C_{l_1,\ldots,l_{n-1}}(l)$. Thus the left member is the number of linearly independent forms of class l apolar to $f_1,\cdots,f_{n-1}$. We suppose that this linear system of dimension L_n-1 of forms of class l is exhibited in terms of the coefficients of $f_1,\cdots,f_{n-1}$ with L_n arbitrary linear homogeneous parameters and apply to the system the condition of apolarity with f_n. Since this yields just enough conditions to determine the L_n homogeneous parameters only L_n-1 of the new conditions can be linearly independent. Each being linear in the coefficients of f_n, the choice of L_n-1 independent conditions yields an apolar form of degree L_n-1 in the coefficients of f_n. Thus the G. C. D. of (7) has a degree in the coefficients of f_i which is *at most* L_i-1.

We proceed to exhibit a more specific method for setting up the apolar form with the actual degree L_n-1 in the coefficients of f_n. This is a consequence of the theorem:

(10) *If, in S_{n-1}, $n-1$ primals, $(\alpha_1 y)^{l_1}=0,\cdots,(\alpha_{n-1}y)^{l_{n-1}}=0$, meet in L_n distinct points, $(p_i\eta)=0$, there is in general no linear identity connecting the k-th powers of these points if $k \geqq l_1+\cdots+l_{n-1}-(n-1)=l-l_n+1$.*

If k were greater than $l-l_n+1$ and an identity for these k-th powers existed, then, by polarizing it, we would have an identity among the lower powers. Thus it will be sufficient to prove the theorem for $k=l-l_n+1$. For this value of k any one of these powers would be apolar to $f_1,\cdots,f_{n-1}$ and we have just seen that the number of linearly independent forms of this class is precisely L_n. That the $(l-l_n+1)$-th powers of these L_n points are themselves in general linearly independent can be proved from their independence in a particular case. This case is the following:

(11) $$(\alpha_1 y)^{l_1}=y_1^{l_1}-y_n^{l_1},\cdots,(\alpha_{n-1}y)^{l_{n-1}}=y_{n-1}^{l_{n-1}}-y_n^{l_{n-1}}.$$

For this special case the points $(p_i\eta)$ are

$$\epsilon_1^{i_1}\eta_1+\cdots+\epsilon_{n-1}^{i_{n-1}}\eta_{n-1}+\eta_n=0,$$

where $\epsilon_1,\cdots,\epsilon_{n-1}$ are primitive roots of unity of indices $l_1,\cdots,l_{n-1}$ respectively. We wish to prove that an identity of the form,

$$\sum_{n-1}\lambda_{i_1 i_2\ldots i_{n-1}}(\epsilon_1^{i_1}\eta_1+\cdots+\epsilon_{n-1}^{i_{n-1}}\eta_{n-1}+\eta_n)^{l_1+\cdots+l_{n-1}-(n-1)}\equiv 0,$$

where the summations for each i_j runs from $i_j=0$ to $i_j=l_j-1$, implies the vanishing of the L_n coefficients λ. This identity yields the following L_n equations in λ:

(12) $$\sum_{n-1}\lambda_{i_1\ldots i_{n-1}}\epsilon_1^{r_1 i_1}\epsilon_2^{r_2 i_2}\cdots\epsilon_{n-1}^{r_{n-1}i_{n-1}}=0$$

$$[r_1+\cdots+r_{n-1}\leqq l_1+\cdots+l_{n-1}-(n-1);\ 0\leqq r_j\leqq l_j].$$

Consider a fixed value k_{n-1} of i_{n-1} and a fixed value s_{n-1} of r_{n-1}. Then (12) can be written in the form

(13) $$\sum_{k_{n-1}=0}^{\Sigma k_{m-1}=l_{n-1}}\sum_{n-2}\lambda_{i_1\ldots i_{n-2}k_{n-1}}\epsilon_1^{r_1 i_1}\epsilon_2^{r_2 i_2}\cdots\epsilon_{n-2}^{r_{n-2}i_{n-2}}\epsilon_{n-2}^{r_{n-2}i_{n-2}}\epsilon_{n-1}^{s_{n-1}k_{n-1}}=0.$$

For fixed values of $i_1,\cdots,i_{n-2}$ and $r_1,\cdots,r_{n-2}$, as s_{n-1} changes from 0 to $l_{n-1}-1$ we have in (13) l_n equations in the l_n expressions

$$\sum_{n-2}\lambda_{i_1 i_2\ldots i_{n-2}k_{n-1}}\epsilon_1^{r_1 i_1}\cdots\epsilon_{n-2}^{r_{n-2}i_{n-2}}$$

for which k_{n-1} varies from 0 to $l_{n-1}-1$. The determinant of this system is $|\epsilon_{n-1}^{ij}|\neq 0$. Hence each of these expressions vanishes and the system (12)

of L_n equations in L_n variables may be replaced by the l_{n-1} systems of L_n/l_{n-1} equations in L_n/l_{n-1} variables expressed by

$$(14) \qquad \sum_{n-2} \lambda_{i_1 \ldots i_{n-2}k_{n-1}} \epsilon_1^{r_1 i_1} \cdots \epsilon_{n-2}^{r_{n-2} i_{n-2}} = 0.$$

In one of these systems defined by fixing k_{n-1} the summation is for $i_1, \cdots, i_{n-2}$ only, and the L_n/l_{n-1} equations are those for which $r_1 + \cdots + r_{n-2} = l_1 + \cdots + l_{n-2} - (n-2)$ since the value $r_{n-1} = l_{n-1} - 1$ has been used. If now we fix i_{n-2} at k_{n-2}, the same argument shows that

$$\sum_{n-3} \lambda_{i_1 \ldots i_{n-3}k_{n-2}k_{n-1}} \epsilon_1^{r_1 i_1} \cdots \epsilon_{n-3}^{r_{n-3} i_{n-3}} = 0,$$

and finally we find that $\lambda_{k_1 \ldots k_{n-1}} = 0$ for any value of $k_1, \cdots, k_{n-1}$ within the limits of $r_1, \cdots, r_{n-1}$.

We supplement the theorem (10) by the following:

(15) *There is in general precisely one linear identity connecting the $(l - l_n)$-th powers of the L_n points of Theorem* (10).

These powers, $(p_i\eta)^{l-l_n}$, are forms to which $f_1, \cdots, f_{n-1}$ are apolar whence only $L_n - 1$ of them can be linearly independent. For, in the identity (9) with $m = -1$, the first term on the left is the number of linearly independent forms of class $l - l_n$, the last term on the left is $+1$, and the intermediate terms are $-C_{l_1, \ldots, n-1}(l - l_n)$. That in general there is only one such identity is clear from the above special case (11). For, an assumed identity would yield equations like (12), with however the limitation on the r's that

$$r_1 + \cdots + r_{n-1} \leq l_1 + \cdots + l_{n-1} - n,$$

which excludes the extreme case

$$r_1 = l_1 - 1,\ r_2 = l_2 - 1, \cdots, r_{n-1} = l_{n-1} - 1.$$

Thus one, and only one, of the equations like (12) would be missing. Since the rank of the system (12) is L_n the rank of the present system is $L_n - 1$, whence the solution is unique.

Since, according to (10), there is no linear identity connecting the powers, $(p_1\eta)^l$, of the L_n points, these powers may be chosen as the L_n linearly independent forms of class l apolar to $f_1, \cdots, f_{n-1}$. The unique apolar form must then be expressible as

$$A = (a\eta)^l = \mu_1(p_1\eta)^l + \cdots + \mu_{L_n}(p_{L_n}\eta)^l.$$

Since A is also apolar to $f_n = (\alpha_n y)^{l_n}$,

$$\mu_1(\alpha_n p_1)^{l_n}(p_1\eta)^{l-l_n} + \cdots + \mu_{L_n}(\alpha_n p_{L_n})^{l_n}(p_{L_n}\eta)^{l-l_n} \equiv 0.$$

However, according to (15), there is just one identity connecting these powers and its coefficients ν_i can of course depend only on $f_1, \cdots, f_{n-1}$. Hence $\mu_i(\alpha_n p_i)^{l_n} = \nu_i$. Thus

(16) $$A = (a\eta)^l = \sum_{i=1}^{i=L_n} [\nu_i (p_i\eta)^l / (\alpha_n p_i)^{l_n}] \cdot \prod_{i=1}^{i=L_n} (\alpha_n p_i)^{l_n}.$$

(17) *The formula* (16) *furnishes the unique apolar form of n generic forms with a degree $L_n - 1$ in the coefficients of f_n, and in terms of constants ν_i, p_{ij} which are determined by $f_1, \cdots, f_{n-1}$ alone.*

Another consequence of the formula (16) is that

(18) *If the n forms $f_1, \cdots, f_n$ are generic except for a common zero at $(p_1\eta) = 0$, then $A = (a\eta)^l = c \cdot (p_1\eta)^l$.*

If this common point exists, it is also on the jacobian $J = (\gamma y)^l$ of the n forms. Thus $(\gamma p_1)^l = 0$, or $(a\eta)^l$ and $(\gamma y)^l$ are apolar. Hence the apolarity invariant $(a\gamma)^l$ vanishes if the n forms have a common zero. This invariant does not vanish identically. Indeed, in the special case (6),

$$(\gamma y)^l = y_1^{l_1-1} \cdots y_n^{l_n-1}, \quad (a\eta)^l = \eta_1^{l_1-1} \cdots \eta^{l_n-1}$$

and these are not apolar. Thus this non-vanishing apolarity invariant of degree L_i in the coefficients of f_i vanishes when the irreducible resultant R of the same degrees vanishes, whence

(19) *The apolarity invariant of the unique apolar form, and of the jacobian, of n forms is the resultant R of the n forms.*

In order to make the numerical constants in A, J, R precise we will suppose that, *in forming the apolarity invariant* $[J, A]$, the form A of class l in $\eta_1, \cdots, \eta_n$ is converted into a differential operator in $\partial/\partial y_1, \cdots, \partial/\partial y_n$ which is applied to J. Then

If the numerical factors of proportionality in A, J and R are so chosen that, when the n given forms are

$$f_1 = a_1 y_1^{l_1} + \cdots, \cdots, f_n = a_n y_n^{l_n} + \cdots$$
$$J = \prod_{i=1}^{i=n} (a_i y_i^{l_i-1}) + \cdots, \quad A = \prod_{i=n}^{i=1} (a_i^{L_i-1} \eta_i^{l_i-1}) + \cdots,$$
$$R = \prod_{i=1}^{i=n} [(l_i - 1)!\, a_i^{L_i}] + \cdots,$$

then $[J, A]$ as defined above is R.

Indeed, when the given forms contain only their leading terms, then A, J, and R contain only their leading terms for which R satisfies the

definition ((19) *et seq.*). Since these terms occur also in the generic case, they serve to define A, J, and R uniquely.

An example of some of the above properties of A is furnished by the case of two binary forms. Let

$$f_1 = (\alpha y)^m = (\alpha_1 y) \cdot (\alpha_2 y) \cdots (\alpha_m y),$$
$$f_2 = (\beta y)^n = (\beta_1 y) \cdot (\beta_2 y) \cdots (\beta_n y).$$

Then, as is well known, there is, when the $m+n$ linear factors are distinct, an identity connecting their $(m+n-2)$-th powers of the form,

$$\sum_{i=1}^{i=m} \lambda_i (\alpha_i y)^{m+n-2} + \sum_{j=1}^{j=n} \mu_j (\beta_j y)^{m+n-2} \equiv 0,$$

where the coefficients λ_i, μ_j are, to within sign, the products of the bilinear invariants of every pair of the linear factors other than that one to which the coefficient is attached. Therefore the unique apolar form is

$$A = \sum_{i=1}^{i=m} \lambda_i (\alpha_i y)^{m+n-2} \equiv -\sum_{j=1}^{j=n} \mu_j (\beta_j y)^{m+n-2}.$$

The first form of A will have, in the coefficients λ_i, a common factor $\Pi(\beta_j \beta'_j)$ which may be discarded. A particular coefficient λ_i will have the factor $(\beta\alpha_1)^n \cdots (\beta\alpha_{i-1})^n \cdot (\beta\alpha_{i+1})^n \cdots (\beta\alpha_m)^n$, and thus, according to (17), will have the degree $m-1$ in the coefficients of A. However, this particular form of A can not be rationalized in terms of the coefficients of f_1. Thus it is to be contrasted with the expressions for A obtained by bordering resultants derived from $J(x^h, y^k)$, these being rational, integral, and of the proper degrees, in the coefficients of both f_1 and f_2.

3. The expression (1) when *m* > *l*. We turn now to the main purpose of this paper and recall a very general identity proved in the memoir cited of Morley-Coble [cf. p. 481 (30)] which reads as follows:

$$\begin{aligned}(21)\quad & J_k(\alpha_1^{l_1}, \cdots, \alpha_n^{l_n}) \cdot (\delta x)^m \\ & \qquad + \sum_{i=1}^{i=n} J_k(\alpha_1^{l_1}, \cdots, \alpha_{i-1}^{l_{i-1}}, \alpha_{i+1}^{l_{i+1}}, \cdots, \alpha_n^{l_n}, \delta^m) \cdot (\alpha_i x)^{l_i} \\ & \equiv J_{k-m}(\alpha_1^{l_1}, \cdots, \alpha_n^{l_n}) \cdot (\delta y)^m \\ & \qquad + \sum_{i=1}^{i=n} J_{k-l_i}(\alpha_1^{l_1}, \cdots, \alpha_{i-1}^{l_{i-1}}, \alpha_{i+1}^{l_{i+1}}, \cdots, \alpha_n^{l_n}, \delta^m) \cdot (\alpha_i y)^{l_i}.\end{aligned}$$

In this the definition of J_k is

$$(22)\quad J_k(\alpha_1{}^{l_1},\cdots,\alpha_n{}^{l_n})$$
$$=\sum(\alpha_1\alpha_2\cdots\alpha_n)(\alpha_1y)^{k_1}\cdots(\alpha_ny)^{k_n}(\alpha_1x)^{l_1-k_1-1}\cdots(\alpha_nx)^{l_n-k_n-1}$$
$$(k_1+k_2+\cdots+k_n=k,\qquad 0\geq k\geq l).$$

Thus J_k has the order k in y and the order $l-k$ in x. Evidently $J_k\equiv 0$ if $k>l$.

The identity (21) is proved by repeated use of the elementary identity connecting $n+1$ linear forms. It has the order k in y, the order $l+m-k$ in x, and it is linear in the coefficients of the given form $(\alpha_iy)^{l_i}$ $(i=1,\cdots,n)$ and of $(\delta y)^m$. Naturally both in (21) and in (22) symbolic terms with negative exponents are not used.

Let, in (21), $k=m>l$. Then the first term of the identity does not appear since, as noted above, $J_k\equiv 0$. The order in x of the identity is l. We take then the apolarity invariant of the terms of the identity in x with the unique apolar form, $A=(a\xi)^l$, the ξ's being contragredient to the x's. Then the remaining terms on the left of (21) disappear, since $(\alpha_ix)^{l_i}$ is apolar to $(a\xi)^l$. The first term on the right of (21) is J in variables x and, according to (20) the apolarity invariant of this with $(a\xi)^l$ is R. Transposing the remaining terms on the right of (21), we have the theorem:

(23) *If $(\delta y)^m$ is an arbitrary form of order $m>l$, then $R\cdot(\delta y)^m$ can be expressed in terms of $(\alpha_iy)^{l_i}$ $(=1,\cdots,n)$ as in (1) with*

$$(\beta_iy)^{m-l_i}=(-1)^{n-i}[J_{m-l_i}(\alpha_1{}^{l_1},\cdots,\alpha_{i-1}{}^{l_{i-1}},\ \alpha_{i+1}{}^{l_{i+1}},\cdots,\alpha_n{}^{l_n},\delta^m),(a\xi)^l],$$

the bracket being the apolarity invariant with respect to x, ξ.

This fundamental expression for $R\cdot(\delta y)^m$ $(m>l)$ has all the elasticity which is desirable in that as $(\delta y)^m$ varies in a pencil, the coefficients $(\beta_iy)^{m-l_i}$ also vary in projective pencils. The coefficients have the desired covariant character but the geometric conditions on them which limit them to the above specific forms are by no means apparent. In the preceding memoir unique coefficients were also obtained by imposing relatively simple geometric conditions but this advantage entailed the introduction of extraneous parameters.

4. Extension of the expression (1) to cover cases $m\leq l$. We again take $k=m$, and again take the apolarity invariant of the identity (21) in x with respect to $(a\xi)^l$. The effect on the terms on the right of (21) is the same as before, as well as the effect on those terms on the left of (21) which are in $\sum_{i=1}^{i=n}$. However, $J_m(\alpha_1{}^{l_1},\cdots,\alpha_n{}^{l_n})$ is no longer identically zero. But the

apolarity invariant of $J_m \cdot (\delta x)^m$ and $(\alpha\xi)^l$ is the apolarity invariant of J_m itself and the polar $(a\delta)^m(\alpha\xi)^{l-m}$ times $[l]_{l-m}\cdot m\,!$. Hence

(24) *If $(\delta y)^m$ is an arbitrary form of order $m \leq l$, then $R\cdot(\delta y)^m$ can be expressed precisely as in* (23) *except for an additive term,*

$$[l]_{l-m}\cdot m\,!\,[J_m(\alpha_1{}^{l_1},\cdots,\alpha_n{}^{l_n}),\,(\alpha\delta)^m(a\xi)^{l-m}].$$

For example, if $m = l$, the extension of the given module by the jacobian of the given forms, gives enough freedom to express all forms of order l.

The unique apolar form, when $R \neq 0$, may be defined as the *fundamental combinant* of the module determined by $(\alpha_1 y)^{l_1},\cdots,(\alpha_n y)^{l_n}$. Indeed, with $l_1 \leq l_2 \leq \cdots \leq l_n$, and with $(a\eta)^l$ given, $(\alpha\, y)^{l_1}$ is the form of lowest order apolar to $(a\eta)^l$ unless $l_1 = l_2 = \cdots = l_k$ in which case any k linearly independent forms of order l_1 apolar to $(a\eta)^l$ constitute the first k forms of the module. Then $(\alpha_{k+1}y)^{l_{k+1}}$ is one form, or one of j linearly independent forms, of order l_{k+1} both apolar to $(a\eta)^l$ and also not found in the module determined by $(\alpha_1 y)^{l_1},\cdots,(\alpha_k y)^{l_k}$. Proceeding in this way the module may be obtained from the apolar forms of $(a\eta)^l$ alone.

To all of the above there is one exception even when $R \neq 0$. This is the case where $l_n > l$ or $0 > l_1 + \cdots + l_{n-1} - n$ so that $l_1 = l_2 = \cdots = l_{n-1} = 1$. Then the apolarity of the last form with $(a\eta)^l$ presents no conditions and there is no significant apolar form of class l when $R \neq 0$. This exception occurs first in the binary case $l_1, l_2 = 1, l_2$ and it persists as new variables and linear forms are added.

As an example of the transition from a module to its unique apolar form and vice versa consider the ternary module determined by a ternary quadratic, cubic, and quartic, say $(\alpha y)^2$, $(\beta y)^3$, $(\gamma y)^4$, each generic. Then, if $(p_1\eta),\cdots,(p_6\eta)$ are the six points common to $(\alpha y)^2$, $(\beta y)^3$, according to (15) and (17), we have

$$(25,\mathrm{a})\qquad A = \lambda_1(p_1\eta)^6 + \lambda_2(p_2\eta)^6 + \cdots + \lambda_6(p_6\eta)^6,$$
$$(25,\mathrm{b})\qquad 0 \equiv (p_1\eta)^2 + (p_2\eta)^2 + \cdots + (p_6\eta)^2.$$

With A as given in (25, a), the conic $(\alpha y)^2$ is the unique conic on $p_1,\cdots,p_6$ whose existence is a consequence of (25, b). The cubic $(\beta y)^3$ is any cubic on $p_1,\cdots,p_6$ which does not contain the conic as a part. The quartic $(\gamma y)^4$ is any quartic apolar to A which is not in the module determined by $(\alpha y)^2$, $(\beta y)^3$. The module depends projectively on the 11 constants involved in the choice of the six points on a conic and the 5 additional constants which arise from the ratios of the λ's. Thus A itself is a quite special sextic of class six subject to 11 conditions.

UNIVERSITY OF ILLINOIS.

ASYMPTOTIC INTEGRATION CONSTANTS.*

By Aurel Wintner.

Let x and f be vectors with a common number of real components, x_i and f_i $(i = 1, \cdots, k)$, and let $f = f(x, t)$ be defined and continuous at every point (x, t) of the product space of the whole Euclidean x-space and of the half-line $0 \leqq t < \infty$. Suppose further that, if the sign of absolute value refers to Euclidean length, the inequality

$$(1) \qquad | f(x, t) | \leqq \lambda(t) \phi(| x |)$$

holds for a positive and, for instance, continuous function $\lambda(t), 0 \leqq t < \infty$, satisfying

$$(2) \qquad \int_0^\infty \lambda(t) \, dt < \infty$$

and for a function $\phi(r)$ which is defined on the closed half-line $0 \leqq r < \infty$, is positive and continuous on the open half-line $0 < r < \infty$ and satisfies

$$(3) \qquad \int_1^\infty (dr)/\phi(r) = \infty.$$

Let the set of all these conditions be denoted by (*).

Clearly, (*) is a restriction of $f(x, t)$ itself (for large $| x | - | t |$), rather than one of the Lipschitz-Osgood difference

$$(4) \qquad f(x^*, t) - f(x^{**}, t)$$

(for small $| x^* - x^{**} |$). Correspondingly, the uniqueness of the solution $x = x(t)$ of

$$(5) \qquad x' = f(x, t), \qquad x(0) = x^0,$$

where the initial vector x^0 is given and x' denotes dx/dt, is not ensured by (*). Nevertheless, it is easy to prove the following lemma (cf. [2]):

(i) *If $f(x, t)$ satisfies (*), then every solution $x = x(t)$ of (5), where x^0 is given arbitrarily, exists for $0 < t < \infty$ and tends to a finite limit vector, $x(\infty)$, as $t \to \infty$.*

* Received April 23, 1946.

Let $(?)$ denote the set of those points of the x-space which represent limiting positions, $x(\infty)$, of solutions, $x = x(t)$, of $x' = f(x, t)$. Then (i) does not claim that $(?)$ is the whole x-space, i. e., that, *instead of the initial vector,* $x(0)$, *one can assign the final vector,* $x(\infty)$, *as an arbitrary integration constant.* In fact, the chances for the truth of this converse of (i) seem to be lowered by the circumstance that the assumption, $(*)$, of (i) does not ensure the uniqueness of the solution, $x = x(t)$, of (5). But it turns out that this impression is misleading, since $(*)$ alone implies that $(?)$ is the whole x-space:

(ii) *If* $f(x, t)$ *satisfies* $(*)$, *then*

$$x' = f(x, t), \qquad x(\infty) = {}^0x, \tag{6}$$

where ${}^0x$ *is given arbitrarily, has a solution* $x = x(t)$, $0 \leqq t < \infty$.

Clearly, the assumptions imposed by $(*)$ on $f(x, t)$ remain satisfied if $f(x, t)$ is replaced by $f(x - a, t)$, where a is an arbitrary constant vector. Hence, it is sufficient to prove (ii) for the case ${}^0x = 0$, where 0 denotes the zero vector. In other words, the assertion of (ii) is that, if $f(x, t)$ satisfies $(*)$, then

$$x(t) \to 0 \text{ as } t \to \infty \tag{7}$$

must hold for some solution $x = x(t)$, $0 \leqq t < \infty$, of $x' = f(x, t)$.

In order to establish the existence of such a solution of $x' = f(x, t)$, choose an arbitrary $t_0 \geqq 0$ and consider a solution $x = x(t)$ satisfying

$$x(t_0) = 0. \tag{8}$$

It is clear from the first of the assertions of (i) that this solution must exist for $0 \leqq t < \infty$. In view of (1), it satisfies the inequality

$$\int_u^v |dx(t)|/\phi(|x(t)|) \leqq \int_u^v \lambda(t)\,dt \tag{9}$$

on every interval $u \leqq t \leqq v$. But the assumptions imposed by $(*)$ on $\phi(r)$, where $0 \leqq r < \infty$, are (1), (3) and the condition that $\phi(r)$ be positive and

continuous when $r \neq 0$, and all these assumptions remain satisfied if $\phi(r)$ is replaced by the function which is max $(\phi(1), \phi(r))$ or $\phi(r)$ according as $0 \leqq r \leqq 1$ or $r > 1$. Hence, it can be assumed that $\phi(r)$ is positive and continuous at $r = 0$ also. Then, since $x(t)$ is differentiable, $1/\phi(|x(t)|)$ is continuous throughout and cannot, therefore, omit any value between values attained. Consequently, from (9) and (8),

$$(10) \qquad \int_0^R (dr)/\phi(r) \leqq \int_0^\infty \lambda(t)\,dt, \text{ where } R = \underset{0 \leqq t < \infty}{\text{l. u. b.}} |x(t)|,$$

the lower limit of integration on the left of (10) being

$$(11) \qquad 0 = x(t_0) = \underset{0 \leqq t < \infty}{\text{g. l. b.}} |x(t))|.$$

The relations (10), (3), (2) imply the existence of a constant C having the property that

$$(12) \qquad |x(t)| < C \text{ for } 0 \leqq t < \infty,$$

where $x(t)$ is any solution satisfying (8) for some t_0, and C is independent of the choice of this solution (including the choice of t_0). However, since $\phi(r)$ has a positive minimum on every closed, bounded interval $0 \leqq r \leqq C$, it is seen from (12) and (9) that

$$(13) \qquad \int_u^v |dx(t)| \leqq c \int_u^v \lambda(t)\,dt,$$

where $c = c(C)$ is independent of the choice of u, v in (13) and of that of t_0 in (8).

Let $t_0 = n(=1, 2, \cdots)$, and let $x^n(t)$, where $0 \leqq t < \infty$, be a corresponding solution of $x' = f(x, t)$. Then, from (13) and (2),

$$(14) \qquad \int_0^\infty |dx^n(t)| < \text{const.}$$

In view of $x^n(n) = 0$, this is a refinement of

$$(15) \qquad |x^n(t)| < C;$$

cf. (8). In addition, (13) and (2) imply that, for every non-negative T,

$$(16)\qquad \int_T^\infty |\, dx^n(t)\,| < \epsilon(T),$$

where $\epsilon = \epsilon(T)$ is independent of n and tends to 0 as $T \to \infty$. Finally, (16) implies, besides the existence of a finite limit vector $x^n(\infty)$, the inequality

$$|\, x^n(T) - x^n(\infty)\,| < \epsilon(T).$$

This means that

$$(17)\qquad x^n(t) \to x^n(\infty) \text{ as } t \to \infty \text{ holds uniformly in } n.$$

Since $f(x, t)$ is continuous on the product space of the whole x-space and of the half-line $0 \leqq t < \infty$, it is seen from (15) and (1) that the sequence $f(x^1(t), t), f(x^2(t), t), \cdots$ is uniformly bounded on every bounded interval of this half-line. But the n-th element of the sequence is the derivative of $x^n(t)$. Hence, the sequence $x^1(t), x^2(t), \cdots$ is equicontinuous on every bounded interval of the half-line. Since it is uniformly bounded, it contains a subsequence which is uniformly convergent on every bounded interval. In view of (15) and of the existence of every $x^n(\infty)$, this subsequence of the functions $x^1(t), x^2(t), \cdots$ can be chosen so as to make the corresponding subsequence of the limits $x^1(\infty), x^2(\infty), \cdots$ a convergent sequence of constants. Then it is seen from (17) that the resulting subsequence of $x^1(t), x^2(t), \cdots$ is uniformly convergent on the half-line $0 \leqq t < \infty$.

Let $x(t)$, where $0 \leqq t < \infty$, denote the limit function of such a subsequence. Then $x = x(t)$ is a solution of $x' = f(x, t)$, since every $x = x^n(t)$ is. Furthermore, $x(\infty)$ exists, by (i) or (14). But if the n-th element of the subsequence is denoted by $x^m(t)$, where $m = m_n$, then

$$\underset{0 \leqq t < \infty}{\text{l. u. b.}} |\, x^m(t) - x(t)\,| \to 0 \text{ as } n \to \infty.$$

In particular, since $x^m(\infty)$ and $x(\infty)$ exist,

$$x^m(\infty) \to x(\infty) \text{ as } n \to \infty.$$

Finally, since $x^n(n) = 0$,

$$x^m(m) = 0 \text{ where } m \to \infty \text{ as } n \to \infty.$$

In view of (17), the last three formula lines imply that the value of the constant $x(\infty)$ cannot be distinct from 0. Since this means that the solution $x(t)$ constructed satisfies (7), the proof of (ii) is complete.

Appendix.

Condition (*) has a dual which, when substituted for (*), leaves the assertion of (i) unaltered but transforms the situation expressed by (ii) into the complete opposite of that situation, in the following sense: The limiting vector $x(\infty)$, instead of being connected with the initial vector $x(0)$ in such a way as to be capable of *every* position in the x-space, becomes a position which is independent of $x(0)$; so that all solution paths of $x' = f(x, t)$ become "confluent" at a single point of the x-space, as $t \to \infty$.

Let (* bis) denote the set of conditions which results from the set (*) if (1) is replaced by

(1 bis) $$x \cdot f(x, t) \leqq -\lambda(t)\phi(x \cdot x)$$

and (2) by

(2 bis) $$\int_0^\infty \lambda(t)\,dt = \infty, \qquad (\lambda \geqq 0)$$

finally (3) by

(3 bis) $$\int_\epsilon^R (dr)/\phi(r) < \infty \text{ if } 0 < \epsilon < R < \infty \qquad (\phi \geqq 0)$$

(it is understood that $x \cdot y$ in (1 bis) denotes the scalar product $x_1 y_1 + \cdots + x_k y_k$). It will be shown that, corresponding to the duals (1 bis), (2 bis), (3 bis) of (1), (2), (3), the "confluent" counterpart of (i) and (ii) can be formulated as follows:

If $f(x, t)$ satisfies (bis), then every solution $x = x(t)$ of (5), where x^0 is arbitrary, exists for $0 \leqq t < \infty$ and satisfies (7).*

First, if $r = r(t)$ denotes $|x|^2$, where $x = x(t)$, then $x' = f(x, t)$ implies that $\frac{1}{2}r'$ is the scalar product $x \cdot f(x, t)$. Hence, from (1 bis),

$$r'(t) \leqq -2\lambda(t)\phi(r(t)),$$

and so, since $\lambda \geqq 0$ and $\phi \geqq 0$, the derivative of $|x(t)|^2$ is non-positive throughout. Consequently, if t increases from $t = 0$ onward, $|x(t)|$ cannot increase. Since this precludes the existence of a $t^0 > 0$ satisfying $|x(t)| \to \infty$ as $t \to t^0 - 0$, and since $f(x, t)$ is supposed to be defined and continuous on the product space of the whole x-space and of the half-line $0 \leqq t < \infty$, it follows from a general fact concerning systems $x' = f$ (cf. [1], p. 177), that every solution $x = x(t)$ of (5) exists for $0 \leqq t < \infty$.

Since the derivative of $r(t) = |x(t)|^2$ is non-positive, it is also seen that there exists a finite limit $r(\infty) \geqq 0$. Hence, in order to complete the proof of the last italicized statement, all that remains to be shown is that $r(\infty) > 0$ is impossible. But this can be concluded from the assumptions which have not been used thus far. In fact, if $r(t)$ does not vanish for all t from a certain $t = t^*$ onward, say for $t \geqq 0$, then, by the last formula line, the inequality

$$\int_0^T |dr(t)| / \phi(r(t)) \geqq \int_0^T \lambda(t)\, dt$$

holds for every $T > 0$, since $r(t)$ is monotone and $\phi \geqq 0$, $\lambda \geqq 0$. Consequently, from (2 bis),

$$\int_0^\infty |dr(t)| / \phi(r(t)) = \infty,$$

and so, since $r(t)$ is a continuous, non-increasing function,

$$\int_{r(\infty)}^{r(0)} (dr)/\phi(r) = \infty, \text{ where } 0 \leqq r(\infty) \leqq r(0) < \infty.$$

It follows therefore from (3 bis) that $r(\infty) > 0$ is impossible.

The "o-theorem" just proved has an "O-variant," which in the linear case can be formulated as follows:

If $A(t)$, where $0 \leqq t < \infty$, is a matrix of n times n real-valued continuous functions satisfying the unilateral restriction

$$\limsup_{t\to\infty} \int_0^t \max_{|y|=1} (y \cdot A(s)y)\, ds < \infty,$$

then every solution vector $x = x(t)$ of $x' = A(t)x$ is bounded as $t \to \infty$.

The restriction is unilateral, since the integral, the upper limit of which is required to be distinct from $+\infty$, is allowed to have the lower limit $-\infty$. In fact, the function integrated is the maximum, rather than the maximum of the absolute value, attained on the real unit sphere $y_1^2 + \cdots + y_n^2 = 1$ by the form which results if the two sets of variables are indentified in the bilinear form belonging to the matrix. In other words, the function integrated

is the greatest, rather than the absolutely greatest, characteristic number of the arithmetical mean of the matrix and its transposed matrix.

The proof follows if $x' = A(t)x$ is multiplied by $x = x(t)$. Since $x \cdot x' = |x| \, |x|'$, this shows that $|x(t)| \, |x(t)|'$ does not exceed $|x(t)|^2$ times the maximum of $y \cdot A(t)y$ on the sphere $|y| = 1$. But $|x(t)|$ cannot vanish at any $t = t_0$ unless it vanishes identically, since $x(t) \equiv 0$ is one, hence the only, solution of $x' = A(t)x$ and of the initial condition $x(t_0) = 0$. Hence, if the trivial solution $x(t) \equiv 0$ is excluded, division by $|x(t)|^2$ is allowed, and so the preceding estimate means that $(\log |x(t)|)'$ does not exceed the maximum of $y \cdot A(t)y$ on the sphere $|y| = 1$. It follows, therefore, from the last formula line that $\log |x(t)|$ *is bounded from above*, which proves that $|x(t)|$ is bounded.

REFERENCES.

[1] A. Wintner, "The infinities in the non-local existence problem of ordinary differential equations," *American Journal of Mathematics*, vol. 68 (1946), pp. 173-178.

[2] ———, "An Abelian lemma concerning asymptotic equilibria," *ibid.*, pp. 451-454.

THE MEASURE THEORETIC APPROACH TO DENSITY.*

By R. CREIGHTON BUCK.

1. Introduction. The density of a set of integers $A = \{a_n\}$ is usually defined as $D(A) = \lim n/a_n$. This function has some of the properties of a finitely additive measure on the countable space composed of the positive integers, although it is true that sets A and B may have a density while $A \cup B$ does not. However, it is clear that generalized definitions of density can be given which apply to all sets of integers, and which are in fact true measures. [1, 231] [4].

The present paper is largely devoted to an analysis of the measure defined on the set of positive integers by applying the Carathéodory extension to a simple basic measure; connection with the theory of Jordan content is very close. This also provides a simple model for classical measure theory; since points are to have zero measure, while the space on which the measure is defined is only countable, we must require only finite additivity. It is also clear that in studying the set of integers, we are studying any countable discrete space, for such a space can be mapped onto I.

In Section **2**, we construct the measure μ and the class $\mathcal{D}_\mu$ of measurable sets. In Section **3**, we prove measurability of certain special sets using number theoretic methods. We discuss in Section **4** certain questions related to sequences of sets and prove that the range of μ is precisely the closed unit interval. Section **5** is devoted to an analysis of what we have called 'quasi-progressions.' It is proved that if α is irrational, the set of integers of the form $[\alpha n + \beta]$ intersects every arithmetic progression in an infinite set. In a sense, this is dual to the fact that if α is irrational, the fractional parts of $\alpha n + \beta$ are everywhere dense in the unit interval. In Section **6**, we discuss a number of properties of ordinary density, and in the following section examine certain of its generalizations, using results dealing with regular summability methods. The concluding section deals with the customary dyadic mapping and questions of the relative measure of classes of measurable or densable sets. Several problems are left open.

2. Measure density. Let I denote the set of all positive integers. On the class of arithmetic progressions, we have what we may call a natural

* Received July 1, 1946.

definition of density. If A is $\{an+b\}$ we define its density to be $\Delta(A) = 1/a$; this we take as a starting point. A measure density on I will be a measure defined for a class of subsets of I which is finitely additive and under which a progression A has measure $\Delta(A)$. We shall require that a point have measure zero, so that the same is true for any finite set; it is then clear that altering a finite number of points of a measurable set will not change its measure. A dot placed above the symbol for a relation will be used to indicate that the relation holds modulo the class of finite sets; thus, $A \dot{\subset} B$ means that if a finite set is deleted from both sets, we will have $A \subset B$, while $A \doteq 0$ means that A itself is finite.

DEFINITION A. *$\mathfrak{D}_0$ is the class of all sets $A \subset I$ which are finite unions of arithmetic progressions, or which differ from these by finite sets.*

This class has the following properties:

(A1) If $A \in \mathfrak{D}_0$, then $A' \in \mathfrak{D}_0$, where A' is the complement of A.

(A2) If $A \in \mathfrak{D}_0$, $B \in \mathfrak{D}_0$, then $A \cup B$ and $A \cap B$ belong to $\mathfrak{D}_0$.

(A3) If $A \in \mathfrak{D}_0$ and $A \doteq B$, then $B \in \mathfrak{D}_0$.

The function Δ may now be extended to $\mathfrak{D}_0$.

DEFINITION B. *If A is a progression $\{an+b\}$, then $\Delta(A) = 1/a$; if A is the union of the disjoint progressions $A_1, A_2, \cdots A_r$, then $\Delta(A) = \Delta(A_1) + \Delta(A_2) + \cdots + \Delta(A_r)$; if $\Delta(A)$ is defined and $A \doteq B$, then $\Delta(B) = \Delta(A)$.*

The measure function Δ clearly has the following properties.

(B1) If A and B belong to $\mathfrak{D}_0$ and $A \dot{\subset} B$, then $\Delta(A) \leq \Delta(B)$.

(B2) If A and B belong to $\mathfrak{D}_0$ and $A \cap B \doteq 0$ then

$$\Delta(A \cup B) = \Delta(A) + \Delta(B).$$

(B3) If A and B belong to $\mathfrak{D}_0$ then

$$\Delta(A \cup B) + \Delta(A \cap B) = \Delta(A) + \Delta(B).$$

This last property is obviously implied by (B2).

We now define an outer measure on I in terms of Δ. The statements that follow are left unproved since they follow as in classical measure theory, or in that of Jordan content. [6]

DEFINITION C. *If* $S \subset I$, $\mu(S) = \inf \Delta(A)$ *for* $A \dot{\supset} S$ *and* $A \varepsilon \mathfrak{D}_0$.

(C1) If $S_1 \dot{\subset} S_2$ then $\mu(S_1) \leq \mu(S_2)$.

(C2) $\mu(S_1 \cup S_2) \leq \mu(S_1) + \mu(S_2)$.

(C3) If $A \varepsilon \mathfrak{D}_0$ then $\mu(A) = \Delta(A)$.

DEFINITION D. $\mathfrak{D}_\mu$ *is the class of all sets* S *for which*

$$\mu(S) + \mu(S') = 1.$$

This class is the Carathéodory extension of $\mathfrak{D}_0$ since the definition above is equivalent to either of the following:

(i) S belongs to $\mathfrak{D}_\mu$ if for any set X,

$$\mu(X) = \mu(X \cap S) + \mu(X \cap S').$$

(ii) S beongs to $\mathfrak{D}_\mu$ if, given $\epsilon > 0$, there exist sets A and B in $\mathfrak{D}_0$ with $A \dot{\subset} S \dot{\subset} B$ and $\Delta(B) - \Delta(A) = \Delta(B - A) < \epsilon$.

In this class, μ is a measure density with the usual properties.

(D1) If $S \varepsilon \mathfrak{D}_\mu$, then $S' \varepsilon \mathfrak{D}_\mu$.

(D2) If S_1 and S_2 belong to $\mathfrak{D}_\mu$, then so do $S_1 \cup S_2$ and $S_1 \cap S_2$.

(D3) If S_1 and S_2 are any two sets of $\mathfrak{D}_\mu$, then

$$\mu(S_1 \cup S_2) + \mu(S_1 \cap S_2) = \mu(S_1) + \mu(S_2).$$

Since μ is obtained as the Carathéodory extension of Δ it is in one sense the most natural measure density in I. In the next section we shall discuss the class $\mathfrak{D}_\mu$ and in particular show that it *properly* contains $\mathfrak{D}_0$.

3. Special sets. An immediate consequence of the definition of $\mathfrak{D}_0$ is that if $A \varepsilon \mathfrak{D}_0$ and $\Delta(A) = 0$, then $A \doteq 0$. We begin by showing that $\mathfrak{D}_\mu$ contains infinite sets of measure zero.

THEOREM 1. *Let* P_0 *be a set of primes such that* $\sum_{P_0} 1/p = \infty$. *Let* S *be a set of integers having the property that if* $p \varepsilon P_0$, *no more than a finite number of integers of* S *are divisible by* p. *Then,* $\mu(S) = 0$.

For example, P_0 can be taken as the set of all primes, or all the primes in an arithmetic progression $\{an + b\}$ where $(a, b) = 1$, or the primes p_{λ_n} where $\lambda_n \sim n$ loglog n. In particular, S can be any coprime set such as the

set of all primes or the set $\{2^{2^n}+1\}$. The less inclusive P_0 is, the more inclusive S may be.

Proof. Let λ be a product of primes of P_0, and let A_k be the arithmetic progression $\{\lambda_n + k\}$ for $k = 1, 2, \cdots \lambda$. Every integer of A_k is divisible by (λ, k) and this in turn is either 1, or itself a product of primes of P_0. Consider the set $S \cap A_k$ in case $(\lambda, k) \neq 1$. Each element of this set is divisible by at least one prime of P_0 which is also a divisor of λ. By hypothesis, only a finite number of terms of S are divisible by any one prime of P_0, and hence by any of the finite collection of primes dividing λ. We conclude that the set $S \cap A_k$ is finite. Writing $I \doteq \underset{k\leq\lambda}{U} A_k$, we have $S = \underset{k\leq\lambda}{U} (S \cap A_k) \doteq \underset{k\leq\lambda}{U'} (S \cap A_k)$, where the dash indicates that the union is to be taken only for k with $(\lambda, k) = 1$. Hence we have $S \dot{\subset} \underset{k\leq\lambda}{U'} A_k$ and $\mu(S) \leq \sum_{k\leq\lambda}' \Delta(A_k) = \phi(\lambda)/\lambda$ where, as usual, $\phi(m)$ is Euler's phi function. Let us now choose λ as $\prod_{p_0\leq m} p_0$ where the subscript indicates that we are considering only primes belonging to P_0. Since $\phi(\lambda) = \lambda \prod_{p_0\leq m} (1 - 1/p_0) \leq \lambda \exp\{-\sum_{p_0\leq m} 1/p_0\}$ we have $\mu(S) \leq \exp\{-\sum_{p_0\leq m} 1/p_0\}$; using our assumed property of the set P_0, and letting m increase, we conclude that $\mu(S) = 0$.

Using similar methods we can show that other special sets possess a μ measure. For example, let S be the set $\{n^2\}$. Modulo λ, this set collapses into a finite collection of distinct numbers $r_1, r_2, \cdots r_a$ which we call the *complete* set of quadratic residues modulo λ. This differs from the customary definition in admitting residues not relatively prime to the modulus. We denote the number of these residues by $w(\lambda) = \alpha$. It is clear that

$$S \subset \{\lambda n + r_1\} \cup \{\lambda n + r_2\} \cup \cdots \cup \{\lambda n + r_a\}$$

and that therefore $\mu(S) \leq w(\lambda)/\lambda$ for all λ. We shall show that lim inf $w(\lambda)/\lambda = 0$, and thus that $\mu(S) = 0$. For this, it would be sufficient to know that $w(n)$ is multiplicative, that if p is any prime $w(p) = (p+1)/2$, and that $\prod_1^n (1 + 1/p_k) = o(p_n)$. However, the following more complete result is of interest in itself[1]

THEOREM 2. *The function $w(n)$ has the following properties which suffice to define it completely:*

[1] A partial solution appears in Uspensky, *Elementary Theory of Numbers*, p. 324.

(i) *If* $(a, b) = 1$, *then* $w(ab) = w(a)w(b)$.

(ii) $w(p^n) = p^{n+1}/(2p+2) + \begin{cases} (p+2)/(2p+2) \text{ for } n \text{ even} \\ (2p+1)/(2p+2) \text{ for } n \text{ odd.} \end{cases}$

(iii) $w(2^n) = 2^{n-1}/3 + \begin{cases} 4/3 \text{ if } n \text{ is even} \\ 5/3 \text{ if } n \text{ is odd.} \end{cases}$

Proof. (i) Let $\alpha = w(a)$, $\beta = w(b)$, and let $r_1, r_2, \cdots r_\alpha, s_1, s_2, \cdots s_\beta$ be the complete sets of quadratic residues modulo a and b respectively. Consider the sets of integers $\{r_i + ax\}$ and $\{s_j + by\}$ where $x = 0, 1, \cdots b-1$, $y = 0, 1, \cdots a-1$, $i = 1, 2, \cdots \alpha$ and $j = 1, 2, \cdots \beta$. The numbers in these sets are the quadratic residues modulo a and b respectively that lie between 0 and ab. Their common part is exactly the complete residue set modulo ab. Since $(a, b) = 1$ we can find unique x and y corresponding to any pair i, j, such that $ax - by = s_j - r_i$. For each pair i, j, there is then a quadratic residue modulo ab, and these are all distinct; their number is then $\alpha\beta = w(a)w(b) = w(ab)$.

(ii) Since $x^2 \equiv (-x)^2$, we have $w(p) = 1 + (p-1)/2 = p^2/(2p+2) = (2p+1)/(2p+2)$, and the formula holds for $n = 1$. To compute $w(p^2)$, we observe that the $\phi(p^2)/2$ integers $1, 2, \cdots p-1, p+1, \cdots (p^2-1)/2$, omitting all multiples of p, have distinct squares modulo p^2. The remaining integers, $p, 2p, \cdots (p-1)p, p^2$ have the same square modulo p^2, namely zero. Thus, $w(p^2) = 1 + p(p-1)/2 = p^3/(2p+2) + (p+2)/(2p+2)$, and the formula holds for $n = 2$. For the general case, suppose that x and y are prime to p while $x^2 \equiv y^2 \pmod{p^n}$. Then $(x+y)(x-y) \equiv 0$ and since both factors cannot be divisible by p unless x and y are, we infer that $x \equiv \pm y \pmod{p^n}$. Thus of the $\phi(p^n)$ integers between 1 and p^n not divisible by p, exactly half give rise to incongruent squares. Suppose now that $x = px'$, $y = py'$ while $x^2 \equiv y^2 \pmod{p^n}$. This implies that $x'^2 \equiv y'^2 \pmod{p^{n-2}}$ and the number of incongruent squares obtained is $w(p^{n-2})$. Combining these, we see that $w(p^n) = \phi(p^n)/2 + w(p^{n-2})$; solving this difference equation with initial conditions corresponding to the values of $w(p)$ and $w(p^2)$, we obtain formula (ii).

(iii) The proof of this is quite similar. We readily find that $w(2) = w(4) = 2$. In the general case, since $x^2 \equiv (-x)^2 \equiv (x + 2^{n-1})^2 \pmod{2^n}$, we need only consider the squares $1^2, 2^2, \cdots (2^{n-2})^2$. By an argument similar to that used in the proof of (ii), we find that $w(2^n) = \phi(2^n)/4 + w(2^{n-2})$; solving this, we arrive at formula (iii).

THEOREM 3. *Lim sup* $w(n)/n = 1/2$ *and* $\liminf w(n)/n = 0$.

From the preceeding theorem, we have $w(p^n)/p^n \leq 2^{-1} + p^{-n}$ and $w(2^n)/2^n \leq 6^{-1} + 2^{1-n}$. Hence, if $n = 2^\beta \prod_1^m p_{r_i}{}^{\beta_i}$ with $\beta_i \geq 1$

$$\begin{aligned} w(n)/n &\leq (6^{-1} + 2^{1-\beta})2^{-m} \prod_1^m (1 + 2p_{r_i}{}^{-\beta_i}) \\ &\leq (6^{-1} + 2^{1-\beta})2^{-m} \exp 2 \sum_1^m p_{r_i}{}^{-\beta_i} \\ &\leq (6^{-1} + 2^{1-\beta})2^{-m} m^2 e^2 \end{aligned}$$

since $\sum_1^m p_{r_i}{}^{-\beta_i} \leq \sum_1^m p_i^{-1} \leq \sum_1^m k^{-1} \leq 1 + \log m$. If n tends to infinity along a sequence of integers in such a manner that m, the number of distinct prime factors, also approaches infinity, then the uniform estimate obtained above shows that $w(n)/n$ will approach zero. If n tends to infinity through a sequence of values for which m is bounded, then we will obtain the maximum values of $w(n)/n$ by considering only $n = p^m$. We have $w(p^m)/p^m = p/(2p + 2) + O(1)p^{-m}$; letting m tend to infinity gives us $p/(2p + 2)$, and as p increases, we obtain lim sup $w(n)/n = 1/2$. For the limit inferior, choose $n = \prod_{p \leq m} p^2$; since $w(p^2) = (p^2 - p + 2)/2$, we see that $w(p^2)/p^2$ is not greater than $1/2$ for $p \geq 3$, and therefore $w(n)/n = \prod_{p \leq m} w(p^2)/p^2 \leq 2^{-\pi(m)}$ which approaches zero as m increases.

COROLLARY. *If S is the set $\{n^2\}$, then $\mu(S) = 0$.*

We have constructed infinite sets having μ measure zero. Since the only set having Δ measure zero are finite sets, we have shown that the inclusion $\mathfrak{D}_0 \subset \mathfrak{D}_\mu$ is strict. From these null sets of $\mathfrak{D}_\mu$ we can build other measurable sets; thus, for example, if $A \varepsilon \mathfrak{D}_0$ and $Z \varepsilon \mathfrak{D}_\mu$ with $\mu(Z) = 0$, then $A \cup Z$ and $A - Z$ are in $\mathfrak{D}_\mu$ and have measure $\Delta(A)$. We observe that if Z is infinite both of the sets $A \cup Z$ and $A - Z$ cannot belong to $\mathfrak{D}_0$, since $Z = (A \cup Z) \cap (A - Z)'$. For comparison we recall that the general Lebesgue measurable set may be expressed either as a G_δ minus a set of measure zero, or as an F_σ plus a set of measure zero. [6, p. 80.] We now show that in the μ measure, no set of $\mathfrak{D}_\mu$ can be expressed in both forms.

THEOREM 4. *If $S = A \cup Z = B - W$, where A and B belong to $\mathfrak{D}_0$ and $\mu(Z) = \mu(W) = 0$, then S is also in $\mathfrak{D}_0$.*

First, we have $\mu(S) = \Delta(A) = \Delta(B)$. We may write S as $(A \cup Z) \cap (B - W) = (A \cap B \cap W') \cup (Z \cap [B - W])$. The last term has measure

zero since it is a subset of Z, so that $\mu(S) = \quad (A \cap B \cap W') = \Delta(A \cap B)$. Now, $\Delta(A) = \Delta(A \cap B) + \Delta(A - B)$; substituting, we see that $\Delta(A - B) = 0$ and that therefore $A - B \doteq 0$. Similarly, from the identity $\Delta(B) = \Delta(B \cap A) + \Delta(B - A)$, we conclude that $B - A \doteq 0$. Combining these we have $A \doteq B$, and since $A \subset S \subset B$, $S \doteq A$ and S belongs to $\mathfrak{D}_0$.

A natural question arises: are all the sets of $\mathfrak{D}_\mu$ obtainable in this fashion, as sets of the form $A \cup Z$ or $A - Z$ where $A \varepsilon \mathfrak{D}_0$ and $\mu(Z) = 0$? The answer lies in the fact that since every set of $\mathfrak{D}_0$ is, except for a finite set, the union of a finite number of progressions, the measure of any set having either of these forms must be a rational number. In the next section, we construct sets of $\mathfrak{D}_\mu$ having any given irrational number for its measure.

4. Limit sets. We begin by considering sequences of sets. One might be led by analogy with ordinary measure to conjecture that if $A_1 \subset A_2 \subset \cdots$ is a sequence of sets of $\mathfrak{D}_0$, then the limit set $S = \lim A_n$ must belong to $\mathfrak{D}_\mu$ and have measure $\mu(S) = \lim \Delta(A_n)$. A simple example shows that this is false; take $A_n = \{1, 2, \cdots n\}$. These sets increase, each has measure zero, but $\lim A_n = I$. However, we observe in this example that if $S = \{1\}$, then $A_n \dot{\subset} S$ for all n and $\mu(S) = \lim \Delta(A_n)$. We might therefore be led to conjecture instead that if $A_1 \subset A_2 \subset \cdots$, then there exists a set S such that $A_n \dot{\subset} S$ for all n and $\mu(S) = \lim \Delta(A_n)$. If this were true, the task of constructing a set with preassigned measure would be very simple. That this conjecture too is false is shown by the following theorem.

THEOREM 5. *There exist sets $A_1 \subset A_2 \subset \cdots$ of $\mathfrak{D}_0$ such that* lim $\Delta(A_n) = 1/3$ *while if $A_n \dot{\subset} S$ for all n, then S has outer measure $\mu(S) = 1$.*

We shall prove that if $A_n \dot{\subset} S$ for all n, then S must have an infinite number of terms in common with any arithmetic progression. This will imply that $\mu(S) = 1$. For, suppose that $B \varepsilon \mathfrak{D}_0$ and that $S \dot{\subset} B$ with $\Delta(B) < 1$. Then, B' is infinite and there is a progression E such that $E \subset B'$; consequently $S \cap E \dot{\subset} S \cap B' \dot{\subset} B \cap B' = 0$ and $S \cap E \doteq 0$, contradicting the property of S mentioned above.

The sets A_n will be defined as $\cup_1^n C_m$ where C_m is the progression $\{4^m N + r_m\}$. We proceed to select the integers r_m. Arrange all arithmetic progressions into a countable sequence $E_1, E_2, \cdots$. Choose $r_1 = 1$ giving us $C_1 = A_1 = \{4N + 1\}$. Let E_{λ_1} be the first set E_i not a subset of A_1 and choose r_2 as any integer in $E_{\lambda_1} - A_1$. We see that $C_1 \cap C_2 = 0$ while $A_2 \cap E_i$ is infinite for each $i \leq \lambda_1$. Having chosen $r_1, r_2, \cdots r_n$, we select r_{n+1} as any integer in $E_{\lambda_n} - A_n$ where E_{λ_n} is the first set E_i following $E_{\lambda_{n-1}}$ that is not a

subset of A_n. The set C_{n+1} is disjoint from all previously defined sets C_m; moreover, $A_n \cap E_i$ is non-void, and hence infinite, for all $i \leq \lambda_n$.

Now, suppose that $A_n \dot{\subset} S$ for all n. Let E_i be any progression and suppose that $S \cap E_i \doteq 0$. But this essentially contains the set $A_n \cap E_i$ for each n, and this is infinite for at least one value of n. Finally, since the sets C_m are pairwise disjoint, $\Delta(A_n) = \sum_1^n \Delta(C_m) = \sum_1^n 4^{-m}$ and $\lim \Delta(A_n) = 1/3$.

Taking complements, we can construct a decreasing sequence of sets A_n such that $\lim \Delta(A_n) = 2/3$, while if $S \dot{\subset} A_n$ for all n, S must have inner measure zero. We remark in passing that although the second conjectured theorem fails for the class $\mathcal{D}_\mu$, it is however true for the usual limit density D defined on the class $\mathcal{D}$. This is proved in Section **6**.

Following similar methods, we now proceed to construct a set S of irrational measure. We first define two collections of sets $\{A_m\}$ and $\{B_k\}$ such that $A_m \subset B_k$ for all m and k while $\lim \Delta(A_m) = \lim \Delta(B_k) = \beta$, an irrational number. Then, from these we construct a set S such that $A_m \dot{\subset} S \dot{\subset} B_k$ for all m and k. From (ii), definition D, it is clear that S belongs to $\mathcal{D}_\mu$ and has measure β.

Our approximating sets will be defined in turn by two auxiliary sequences of sets, $\{C_i\}$ and $\{D_j\}$; we set

$$A_m = C_1 \cup C_2 \cup \cdots \cup C_m;$$
$$B_k = C_1 \cup C_2 \cup \cdots \cup C_{k-1} \cup D_k.$$

In order that $A_m \subset B_k$ for all m and k, it is sufficient to have $C_m \subset D_k$ for $k = 2, 3, \cdots m$, and for all m. We will also want the sets C_i to be disjoint. Let C_m be the arithmetic progression $\{2^{m(m+1)/2}N + \gamma_m\}$ and D_k the progression $\{2^{k(k+1)/2-1}N + \gamma_k\}$ where γ_n has yet to be determined. It is easily seen that in order to have the sets C_i disjoint, it is necessary and sufficient that for all m

$$\gamma_m \not\equiv \gamma_k (\text{mod } 2^{k(k+1)/2})$$

for $k = 2, 3, \cdots m-1$. Likewise, in order to have $C_m \subset D_k$ we must have

$$\gamma_m \equiv \gamma_k (\text{mod } 2^{k(k+1)/2-1})$$

for all m and for $k = 2, 3, \cdots m$. Combining these, we must solve the set of congruences

$$\gamma_m \equiv \gamma_k + 2^{k(k+1)/2-1} (\text{mod } 2^{k(k+1)/2}).$$

For initial conditions, we take $\gamma_1 = 0$, $\gamma_2 = 1$. A solution is then given by

$$\gamma_n = 1 + 2^2 + 2^5 + \cdots + 2^{n(n-1)/2-1}.$$

Our sets A_m and B_k become

$$\begin{aligned}
A_m &= \{2N\} \cup \{8N+1\} \cup \{64N+5\} \cup \{1024N+37\} \cup \cdots \cup C_m;\\
B_1 &= \{N\} = I;\\
B_2 &= \{2N\} \cup \{4N+1\};\\
B_3 &= \{2N\} \cup \{8N+1\} \cup \{32N+5\}
\end{aligned}$$

and so on. Since the sets C_i are pairwise disjoint, $\Delta(A_m) = \sum_1^m \Delta(C_i) = \sum_1^m 2^{-i(i+1)/2}$, and as m increases this sum approaches the irrational number β whose dyadic expansion is $.10100100010\cdots$. Computing $\Delta(B_k)$ we have $\Delta(B_k) \leq \sum_1^{k-1} \Delta(C_i) + \Delta(D_k) = \sum_1^{k-1} 2^{-i(i+1)/2} + 2^{1-k(k+1)/2}$ which again approaches β as k increases. Finally, we set $S = UA_m = UC_m$. We certainly have $A_m \subset S \subset B_k$ for all m and k.

This construction can be carried out for any irrational number. If β is expressed in the dyadic form $\sum_1^\infty 2^{-\lambda_n}$, we choose C_m as $\{2^{\lambda_m}N + \gamma_m\}$ where $\gamma_1 = 0$, and $\gamma_{n+1} = (2^{\lambda_1} + 2^{\lambda_2} + \cdots + 2^{\lambda_n})/2$. Then, the set $S = UC_m$ belongs to $\mathfrak{D}_\mu$ and has measure $\mu(S) = \beta$. Recalling that as A ranges over the countable set $\mathfrak{D}_0$, $\mu(A)$ takes on every rational value, we have proved the following result.

THEOREM 6. *The set of values of $\mu(S)$ for S in $\mathfrak{D}_\mu$ is exactly the closed unit interval.*

An obvious consequence of this is that the class $\mathfrak{D}_\mu$ has cardinal number c, the cardinal of the continuum; moreover, the class $\mathfrak{D}_\mu/\mathcal{K}$ where $\mathcal{K}$ is the class of null sets, also has cardinal c. For comparison we recall that while there are 2^c Lebesgue measurable sets, the class of measurable sets modulo the null sets has cardinal c.

5. Quasi-progressions. In this section, $[x]$ will always denote the greatest integer in x, while $((x))$ will be the fractional part of x: $x = [x] + ((x))$. Let $A = \{a_n\}$ where $a_n = [\alpha n + \beta]$. If α is an integer, then $a_n = \alpha n + [\beta]$ and A is an arithmetic progression. If α is rational, then A is a finite union of progressions. If we suppose that $\alpha = p/q$, and write $u = mq + r$, where $1 \leq r \leq q$, we find that

$$A = \{pm + b_1\} \cup \{pm + b_2\} \cup \cdots \cup \{pm + b_q\}$$

where $b_r = [\beta + rp/q]$. The case of α irrational remains; the set A will

then be called a quasi-progression. We assume that $\alpha > 1$. We will show that such a set does not belong to $\mathcal{D}_\mu$ and that moreover $\mu(A) = \mu(A') = 1$ so that A is extremal—i. e., has outer measure one and inner measure zero.

LEMMA 1. *Let A and B be two non-negative integers, $A \geq 2$, and α and β two positive real numbers with $0 \leq \beta < 1$. Suppose that we can find a rational number p/q with $(p,q) = 1$ such that*

(i) $$0 \leq \alpha - p/q \leq q^{-3/2}$$

(ii) $$\sqrt{q} > A(1+\sqrt{3})/(2-2\beta).$$

Then, there exists an integer r, $0 < r \leq A$ and integers n and m with $0 < n \leq Aq$, $0 \leq m < p$ such that

(iii) $$pn - Aqm = Bq + r$$

(iv) $$[\alpha n + \beta] = Am + B.$$

Proof. Part (iii) is immediate; choose r so that $Bq + r$ is divisible by $(p, qA) = (p, A) \leq A$, and then choose n and m as the least positive solutions of $px - Aqy = Bq + r$. Now, put

$$\begin{aligned}\delta &= \{(\alpha n + \beta) - (Am + B)\}/n \\ &= \alpha - p/q + (np - Amq - Bq)/nq + \beta/n.\end{aligned}$$

Using (i) and (iii), we have

$$\begin{aligned}0 \leq \delta &\leq q^{-3/2} + r/nq + \beta/n \\ &\leq \left\{ \frac{A\sqrt{q} + A}{q} + \beta \right\} \frac{1}{n}.\end{aligned}$$

If (ii) holds then

$$\sqrt{q} > \frac{A + \{A^2 + 4A(1-\beta)\}^{\frac{1}{2}}}{2(1-\beta)}$$

and

$$(1-\beta)q - A\sqrt{q} - A > 0$$

so that

$$0 \leq \delta < \{(1-\beta) + \beta\}(1/n) = 1/n.$$

Hence,

$$0 \leq \alpha n + \beta - (Am + B) < 1,$$

and (iv) holds.

LEMMA 2. *If α is irrational and if we can choose an infinite number of distinct rationals p/q for which* (i) *and* (ii) *hold, then there are an infinite number of distinct pairs n, m, for which* (iv) *holds.*

Suppose that as we select p/q, only a finite number of distinct pairs n, m, appear. There must be an infinite number of distinct rationals p/q which correspond to the same pair n, m. Since the denominators of these rationals will increase without bound, (i) implies that $\alpha = \lim p/q$. From (iii), however, we see that $p/q = (Am + B)/n + r/nq$; in this, m and n are fixed and r, although it depends on p and q, always lies between 0 and A. Taking limits, we obtain $\alpha = (Am + B)/n$, a rational number.

Let us now introduce the notation $k(\alpha, \lambda)$ for the greatest lower bound of the numbers $q^{\lambda}(\alpha - p/q)$ taken over all rational numbers p/q, with $(p, q) = 1$, and $p/q < \alpha$. If α is algebraic of degree n, then $k(\alpha, n) > 0$, and if α is rational, $k(\alpha, \lambda) = 0$ for all λ. [3; 157, 160.]

LEMMA 3. *If α is irrational, and $\lambda < 2$ then $k(\alpha, \lambda) = 0$.*

This is well known for λ integral; when $\lambda = 1$, it is equivalent to stating that $\liminf ((n\alpha)) = 0$. For $\lambda < 2$, the usual proof by means of the successive convergents of the continued fraction expansion of α still suffices. [3, Theorem 171.]

THEOREM 7. *If $\alpha > 1$ is irrational, and β is any positive real number, then the set $A = \{a_n\}$ where $a_n = [\alpha n + \beta]$ has an infinite number of terms in common with any arithmetic progression, and has outer measure one, inner measure zero.*

From the lemmas, we can find an infinite number of distinct pairs n, m, for which $[\alpha n + \beta] = Am + B$, given any two integers A and B. In Lemma 1, we required β to lie between 0 and 1. This restriction is inessential, for if $\beta = b + \beta_0$ where $0 \leq \beta < 1$, then $[\alpha n + \beta] = b + [\alpha n + \beta_0]$. The proof that $\mu(A) = 1$ is the same as in Theorem 5. To show that A has inner measure zero, we must show that A contains no infinite set of $\mathfrak{D}_0$. Suppose $\{an + b\} \dot{\subset} A$; there is then a sequence of integers λ_i such that $an + b = [\alpha\lambda + \beta]$ for all large n. Suppose that $\alpha > 2$; the case where $1 < \alpha < 2$ may be treated in much the same way. Taking first differences, we have, again for all large n, $a = \alpha(\lambda_{n+1} - \lambda_n) + \epsilon_n - \epsilon_{n+1}$ where $\epsilon_n = ((\alpha\lambda_n + \beta))$. Set $\lambda_{n-1} - \lambda_n = \theta_n$, integral, and let $\underline{\theta}$ and $\bar{\theta}$ be the minimum and maximum values of θ_n for large values of n. Then, $\alpha\theta_n = a + \epsilon_n - \epsilon_{n+1}$ and since $0 \leq \epsilon_n < 1$, $0 \leq \alpha(\bar{\theta} - \underline{\theta}) < 2$. But, $\alpha > 2$ so that $\underline{\theta} = \bar{\theta}$ and $\theta_n \equiv \theta$ for all large n; λ_n is then $n\theta + \lambda_0$ and $\epsilon_{n+1} - \epsilon_n$ is a constant c, for all large n. This in turn gives us $\epsilon_m = mc + \epsilon_0$ and unless c is zero, these would be unbounded. We have therefore shown that

$\epsilon_n = ((\alpha\lambda_n + \beta)) = ((\theta\alpha n + \alpha\lambda_0 + \beta))$ is constant. This contradicts the familiar assertion that if γ is irrational, the points $((\gamma m + d))$ are everywhere dense on $(0, 1)$. [3, 364]. Since $\mu(A) + \mu(A') = 2 \neq 1$, A is clearly not measurable.

6. Limit density. We now consider a more inclusive class of sets of integers in place of $\mathcal{D}_\mu$. Let $\mathcal{D}$ be the class of all sets $A = \{a_n\}$ where $a_n = \alpha n + o(n)$, for some positive real number $\alpha \geq 1$. This class clearly includes $\mathcal{D}_0$ and, as we shall prove shortly, it strictly includes $\mathcal{D}_\mu$ as well. It is the class with which the term density is usually associated. For an arbitrary set $A \subset I$, we first define two important functions. $\alpha(n)$ will denote the characteristic function of the set A, taking the value 1 for $n \in A$ and 0 for $n \notin A$. $A(n)$, the distribution function of the set A, is defined as $\sum_1^n \alpha(k)$ and is the number of integers in A not greater than n. $\beta(n)$ and $B(n)$ will be the corresponding functions for a set B. The characteristic function of the set $A \cap B$ is $\alpha(n)\beta(n)$, and that of $A \cup B$ is $\alpha(n) + \beta(n) - \alpha(n)\beta(n)$.

The quotient $A(n)/n$ is the density of the set A in the interval $(1, n)$; the number $\lim A(n)/n$ when it exists is denoted by $D(A)$ and is called the density of A. Clearly, the set $\{a_n\}$ where $a_n = \alpha n + o(n)$ has density $D(A) = 1/\alpha$. More generally, the following equivalence is easily shown: *if* $A = \{a_n\}$ *and* $A \in \mathcal{D}$, *then* $D(A) = \lim A(n)/n = \lim n/a_n = \lim (1/n) \sum_1^n \alpha(k)$.

We observe that the last expression is simply the Cesàro limit of the characteristic function $\alpha(n)$. We can therefore produce sets for which density is not defined by selecting any sequence of zeros and ones whose $(C, 1)$ limit fails to exist, and forming the set whose characteristic function it is.

Unfortunately, the class $\mathcal{D}$ is not a measure class and D is not a measure. Some of the measure properties hold; if A belongs to $\mathcal{D}$ so does A', and if A and B belong to $\mathcal{D}$ and are disjoint, then $A \cup B$ belongs to $\mathcal{D}$ and $D(A \cup B) = D(A) + D(B)$. However, if A and B are not disjoint, neither $A \cup B$ nor $A \cap B$ need belong to $\mathcal{D}$; [11, 16]. Consider, for example, the sets defined as follows: let $\Lambda = \{\lambda_n\}$ be a set of integers not having a density, let $A = \{3N\}$ and $B = \{b_n\}$ where $b_n = a_n$ if $n \in \Lambda$, and $b_n = 1 + a_n$ otherwise. Both A and B have density $1/3$, but $A \cap B = \{3\lambda_n\}$ which fails to have a density.

In spite of this, we can still define an outer density, $\omega(S)$ as $\inf D(A)$ taken over all sets A in $\mathcal{D}$ which contains S. Pólya has proved that this

outer density can be analytically expressed, and is in fact the Pólya maximum density; that is,

$$\omega(S) = \bar{D}_1(S) = \lim_{\theta\to 1} \limsup_{n\to\infty} \frac{S(n) - S(\theta n)}{n - \theta n}.$$

More precisely, Pólya proved that if $\bar{D}_1(S) = d$, then there exists a set B of $\mathfrak{D}$ having density d and containing S. [7, 562.] The minimum density $\underline{D}_1(S)$ is obtained if lim sup is replaced by lim inf. Then, $\underline{D}_1(S) + \bar{D}_1(S') = 1$, and if $\underline{D}_1(S) = d$, there is a set A in $\mathfrak{D}$ having density d and contained in S. Thus, Pólya minimum density is inner density, and is $1 - \omega(S')$.

If we attempt to extend $\mathfrak{D}$ in the usual manner by means of this outer density ω, we find that the extension is not proper. Let $\mathfrak{D}_\omega$ be all sets S for which $\omega(S) + \omega(S') = 1$.

THEOREM 8. $\mathfrak{D}_\omega = \mathfrak{D}$.

If S belongs to $\mathfrak{D}_\omega$, there must exist sets A and B in $\mathfrak{D}$ with $A \subset S \subset B$ and $D(A) = D(B) = \omega(S)$. The continued inclusion implies that $A(n) \leq S(n) \leq B(n)$, and dividing by n and taking limits, we find that $\lim S(n)/n = \omega(S)$, and $S \varepsilon \mathfrak{D}$.

Since $\mathfrak{D}_0 \subset \mathfrak{D}$, it is true that $\mathfrak{D}_\mu \subset \mathfrak{D}$. In the previous section, we showed that the quasi-progression $\{[\alpha n + \beta]\}$ did not have a μ measure; since it clearly has density $1/\alpha$ we see that $\mathfrak{D}_\mu \subset \mathfrak{D}$ with strict inclusion. Moreover, $\mu(S) \geq \bar{D}_1(S)$ for all sets S.

Rather more useful than the maximum or minimum densities are the upper and lower densities, defined by:

$$UD(A) = \limsup A(n)/n$$
$$LD(A) = \liminf A(n)/n.$$

It is clear from the definitions that $\underline{D}_1(A) \leq LD(A) \leq UD(A) \leq \bar{D}_1(A)$. All inequality signs may be strict as is shown by the example

$$\alpha(n) = (1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, \cdots)$$

for which $\bar{D}_1(A) = 1$, $UD(A) = 2/3$, $LD(A) = 1/3$, $\underline{D}_1(A) = 0$. If $C = A \cup B$, then $C(n) \leq A(n) + B(n)$ and $C(n) - C(\theta n) = \sum_{n\theta}^{n} \gamma(k) \leq \sum_{n\theta}^{n} \{\alpha(k) + \beta(k)\} = \{A(n) - A(\theta n)\} + \{B(n) - B(\theta n)\}$. Consequently, $UD(A \cup B) \leq UD(A) + UD(B)$ and $\bar{D}_1(A \cup B) \leq \bar{D}_1(A) + \bar{D}_1(B)$.

Suppose that we have a sequence of sets $A_1 \supset A_2 \supset A_3 \supset \cdots$ with $D(A_n) = d$ for all n. As observed in Section **4**, we cannot expect properties

of the sets to carry over to their intersection. For example, if $A_n = \{n, n+1, \cdots\}$, then $\bigcap A_n$ is void, while $D(A_n) = 1$. However, by suitably modifying the notion of the intersection of a family of sets, we can prove a rather strong theorem of this type. It is the same theorem that in Section **4** we proved did *not* hold for $\mathfrak{D}_\mu$. (Theorem 5.)

THEOREM 9. *If $A_1 \supset A_2 \supset A_3 \cdots$ then there exists a set B such that $B \dot{\subset} A_n$ for $n = 1, 2, \cdots$ and such that $LD(B) = \lim LD(A_n)$, $UD(B) = \lim UD(A_n)$. In particular, if $A_n \in \mathfrak{D}$, then $B \in \mathfrak{D}$ and $D(B) = \lim D(A_n)$.*

Proof. Since $\liminf A_k(n)/n = LD(A_k)$, and $\limsup A_k(n)/n = UD(A_k)$, we can choose a sequence of integers $n_1 < n_2 < n_3 \cdots$ such that $A_k(n)/n \geq (1 - (1/k+1))LD(A_k)$ for all $n > n_k$ and $A_k(n)/n \geq (1 - (1/k+1))UD(A_k)$ for $n = 1 + n_k$. That is, n_k is an integer such that the partial density of A_k is not too small beyond it, while it is large enough at least once. Let I_k be the segment composed of all n with $n_k < n \leq n_{k+1}$. We construct the set B by assigning to it all the integers of A_k lying in I_k; thus, $B \cap I_k = A_k \cap I_k$ for all k. We must show that B has all the desired properties. First, since $A_k \subset A_m$ for all $k > m$, and $I_k \doteq 0$,

$$B = \bigcup B \cap I_k = \bigcup (A_k \cap I_k) \dot{\subset} A_m \cap \bigcup_{k \geq m} I_k \dot{\subset} A_m \text{ and therefore } B \dot{\subset} A_m,$$

holding for any m. This implies that $UD(B) \leq \text{g.l.b.}\, UD(A_m) = \lim UD(A_m)$ and $LD(B) \leq \lim UD(A_m)$. We compute $B(n)/n$ and show that we can reverse these inequalities, thus establishing equality. If $n \in I_k$, then

$$B(n) = A_1(n_2) - A_1(n_1) + \cdots + A_{k-1}(n_k) - A_{k-1}(n_{k-1}) + A_k(n) - A_k(n_k).$$

Since $A_{r+1} \subset A_r$, $A_r(m) - A_{r+1}(m) \geq 0$ and

$$B(n) \geq A_k(n) - A_1(n_1).$$

For all n in I_k,

$$B(n)/n \geq A_k(n)/n - o(1) \geq (1 - (1/k+1))LD(A_k) - o(1)$$

and as n increases, $\liminf B(n)/n = LD(B) \geq \lim LD(A_k)$. For the upper density, we observe that there is an n in I_k for which $B(n)/n \geq A_k(n)/n - o(1) \geq (1 - (1/k+1))\, UD\,(A_k) - o(1)$ so that $\limsup B(n)/n = UD(B) \geq \lim UD(A_k)$.

COROLLARY. *If $A_1 \subset A_2 \subset A_3 \subset \cdots$ then there exists a set B such that $A_n \dot{\subset} B$ for all n, while $UD(B) = \lim UD(A_n)$ and $LD(B) = \lim LD(A_n)$.*

The sets B are not unique; however, if two sets B_1 and B_2 satisfy the conditions of the theorem, it is evident that the sets $B_1 - B_2$ and $B_2 - B_1$ will both be of density zero. Thus, if we identify two sets whose symmetric difference is null, these modified intersections of the sets A_n are unique. The identification yields a homomorphic image of the boolean ring of subsets of I whose kernel is the ideal of all sets having zero density. In view of this, it is perhaps important to point out that the set B can be chosen so that $A_n + B$, the symmetric difference, is finite and not merely null.

We now introduce the interval sequence associated with a set A. Suppose that A is formed in the following way: a block composed of the first g_1 integers, followed by a gap of length g_2, and then by a block of g_3 consecutive integers, and then a gap of length g_4, and so on. The sequence of integers $\{g_i\}$ determines the set A uniquely. g_1 may be zero, but all others are positive. Put $G_n = \sum_1^n g_i$.

THEOREM 10. $UD(A) = \limsup V_n/G_{2n-1}$

$LD(A) = \limsup V_n/G_{2n}$

where $V_n = g_1 + g_3 + \cdots + g_{2n-1}$.

The greatest values of $A(m)/m$ occur when m is the last term of a block of consecutive integers, and the least values when m is the last term of a void interval. In the first instance, $m = G_{2n-1}$ and $A(m) = g_1 + g_3 + \cdots + g_{2n-1} = V_n$, giving us the first relation above. For the second, $m = G_{2n}$ and $A(m)$ is again V_n.

THEOREM 11. $UD(A) \geq UD(A) \liminf G_{2n-1}/G_{2n} \geq LD(A)$

$\{1 - LD(A)\} \geq \{1 - LD(A)\} \liminf G_{2n}/G_{2n+1} \geq \{1 - UD(A)\}$.

The left hand sides of these inqualities are obvious since G_n is an increasing sequence. For the right half of the first,

$$\begin{aligned} UD(A) \liminf G_{2n-1}/G_{2n} &= \limsup (V_n/G_{2n-1}) \liminf (G_{2n-1}/G_{2n}) \\ &\geq \liminf V_n/G_{2n} = LD(A). \end{aligned}$$

In the second,

$$\begin{aligned} \{1 - LD(A)\} \liminf G_{2n}/G_{2n+1} &\geq \liminf (1 - V_n/G_{2n})(G_{2n}/G_{2n+1}) \\ &\geq \liminf (G_{2n} - V_n)/G_{2n+} = \liminf (G_{2n+1} - V_{n+1})/G_{2n+1} \\ &\geq 1 - UD(A). \end{aligned}$$

COROLLARY 1. *If $A \varepsilon \mathfrak{D}$ and $0 < D(A) < 1$, then* $\lim G_n/G_{n+1} = 1$.

For, $LD(A) = UD(A) = d$, and $d \neq 0$, $1 - d \neq 0$. Hence, by the theorem, $\liminf G_{2n-1}/G_{2n} = \liminf G_{2n}/G_{2n+1} = 1$; since $G_n/G_{n+1} \leq 1$, the result follows. It is clear that this limit need not exist if $D(A) = 0$ or 1.

COROLLARY 2. *If* $\liminf G_n/G_{n+1} = 0$, *then* $UD(A) = 1$; $LD(A) = 0$.

Far weaker assumptions are sufficient to yield extremal maximum and minimum densities.

THEOREM 12. *If* $\liminf G_n/G_{n+1} < 1$, *then* $\bar{D}_1(A) = 1$, $\underline{D}_1(A) = 0$.

From the hypothesis, $\limsup g_{2n-1}/G_{2n-1} > 0$ and $\limsup g_{2n}/G_{2n} > 0$. Then, for values of θ sufficiently close to 1, g_{2n-1}/G_{2n-1} and g_{2n}/G_{2n} exceed $1 - \theta$ for an infinite number of integers n. Put $N = G_{2n-1}$; since g_{2n-1} is the length of a filled block of integers, and is greater than $N(1 - \theta)$, $A(N) - A(N\theta) = N - N\theta$, so that the quotient of these has the value 1 for infinitely many N. Hence, $\bar{D}_1(A) = 1$. In the same manner, putting $N = G_{2n}$, and recalling that g_{2n} is the length of a gap, we find that $A(N) - A(N\theta) = 0$ infinitely often, and $\underline{D}_1(A) = 0$.

Sets having outer measure 1 and inner measure 0 are called extremal sets. For a limit density, A is extremal if $UD(A) = 1$, and $LD(A) = 0$. The possibility of expressing a space as a sum of disjoint extremal sets has been discussed for ordinary Lebesgue measure [10]. In this connection, the following theorem is of interest.

THEOREM 13. *The set I may be split into a countable collection of disjoint sets which are extremal in the sense of density. That is, $I = \bigcup_1^\infty A_k$ where $A_n \frown A_m = 0$ if $n \neq m$, and $UD(A_k) = 1$, $LD(A_k) = 0$.*

A set that is extremal in the sense of density is also extremal with respect to inner and outer density, and thus certainly extremal for the μ measure discussed in the previous sections of this paper. We construct the sets A_k so as to have very long blocks of consecutive integers, followed by very long gaps. Choose an interval sequence $\{g_i\}$ so that $\lim G_n/G_{n+1} = 0$. For example, take $g_n = n(n!)$. By Corollary 2, Theorem 11, the set corresponding to this will be extremal. We must next split up I around this set. To simplify the notation, we shall write $g(n)$ and $G(n)$ for g_n and G_n. Let $E(n)$ denote the interval of integers k satisfying $G(n-1) < k \leq G(n)$. Choose any ω^2

type ordering of the positive integers, $\lambda_{k,j}$, such that $1+\lambda_{k,j}<\lambda_{k,j+1}$. For example, we could use the following:

$$(\lambda_{k,j}) = \begin{bmatrix} 1 & 3 & 6 & 10 & . & . & . \\ 2 & 5 & 9 & 14 & . & . & . \\ 4 & 8 & 13 & 19 & . & . & . \\ 7 & 12 & . & . & . & . & . \\ . & . & . & . & . & . & . \end{bmatrix}.$$

Then, the sets A_k are given by

$$A_k = \bigcup_{j=1}^{\infty} E(\lambda_{k,j}).$$

Since the intervals $E(n)$ are all disjoint, and the $\lambda_{k,j}$ all distinct, these sets are all disjoint; since the array $(\lambda_{k,j})$ contains all the integers, $\bigcup_1^{\infty} A_k = I$. The terms of A_k fall into blocks of the form $[G(\lambda_{k,j}-1),\ G(\lambda_{kj})]$ and the greatest values of $A_k(n)/n$ must occur when n is a $G(\lambda_{kj})$. Computing this, we have

$$\begin{aligned} A_k(n)/n &= \frac{g(\lambda_{k1})+g(\lambda_{k2})+\cdots+g(\lambda_{kj})}{G(\lambda_{kj})} \\ &\geqq g(\lambda_{kj})/G(\lambda_{kj}) \end{aligned}$$

by virtue of the restriction on the ordering λ_{kj}. Then, since $\lim G_n/G_{n+1}=0$, $\limsup A_k(n)/n \geqq 1$ and $UD(A_k)=1$. The union of two sets having upper density 1 also has upper density 1; since $A_k' = \bigcup_{n\neq k} A_n$, $UD(A_k')=1$ and therefore $LD(A_k)=0$.

7. Generalized limit densities. Perhaps the most apparent method of generalizing the density discussed in the previous section is to replace the operation 'lim' in the definition $D(A)=\lim A(n)/n$ by a generalized limit. We can always get a total measure density—a measure defined for *all* subsets of the space—by writing $D_L(A) = \mathrm{LIM}\, A(n)/n$ where LIM is a limit defined for all bounded sequences [1, 34]. Otherwise expressed, let L be any linear functional on the space of all bounded sequences such that $L(1)=1$, where $1=(1,1,1,1,\cdots)$ and define $\Delta_L(A)$ to be $L(\alpha)$ where $\alpha=(\alpha(n))$ is the characteristic function of the set A [4]. L can here be chosen as an extension of the functional Δ defined over the linear span of the characteristic functions of the sets in $\mathcal{D}_0$ and the resultant measure density will be an extension of Δ to all subsets of I [1, 231].

More useful, perhaps, are the limit densities obtained by replacing

ordinary limit by a regular summability method; let T be a transformation of the space of sequences given by a real or complex matrix (a_{nv}) so that if $x = (x_i)$, and $y = (y_i)$ and $y = T(x)$, then $y_n = \sum_{v=1}^{\infty} a_{nv} x_v$. We define a generalized limit by writing $T\text{-lim}\, x = \lim y$, whenever this ordinary limit exists. T is regular if this limit is an extension of the usual limit; the Toeplitz necessary and sufficient conditions for regularity are [9]

(i) $$\sum_{v=1}^{\infty} |a_{nv}| \leq M, \text{ for all } n$$

(ii) $$\lim_{n\to\infty} a_{nv} = 0, \text{ for all } v$$

(iii) $$T\text{-lim}\,(1) = 1.$$

Instead of taking $D(A)$ to be $\lim A(n)/n$ we express this in terms of the characteristic function and write this as $(C,1)\text{-lim}\,\alpha$. In order to actually have an extension of ordinary density, we shall require that T be regular over $(C,1)$. This is equivalent to requiring that

(iv) $$\sum_{v=1}^{\infty} v\,|a_{nv} - a_{n,v+1}| \leq K, \text{ all } n.$$

We now define our generalized density by $D_T(A) = T\text{-lim}\,\alpha$, and denote the class of sets for which this exists by $\mathfrak{D}_T$. Clearly, $\mathfrak{D} \subset \mathfrak{D}_T$. We define upper and lower T density in a corresponding fashion as $UD_T(A) = T\text{-lim sup}\ \alpha$, and $LD_T(A) = T\text{-lim inf}\,\alpha$, where $T\text{-lim sup}\,x$ is $\limsup T(x)$. Theorems on regular transformations will now give rise to corresponding theorems on generalized densities. We first prove that no generalized limit density can be total.

THEOREM 14. *The class $\mathfrak{D}_T$ does not contain all the subsets of I.*

We must produce a set of integers not in $\mathfrak{D}_T$. This is equivalent to finding a characteristic function $\alpha(n)$ whose T limit does not exist. A classical theorem of Steinhaus asserts that given any regular transformation T, there is a sequence of zeros and ones not summable T [8].

Let B denote all bounded real sequences, C all convergent sequences, N all null sequences, and Z all sequences of zeros and ones. If x and y are two sequences xy will be the sequence $(x_n y_n)$; we write $x \leqq y$ if $x_n \leqq y_n$ for all n.

THEOREM 15. *If $x \varepsilon B$, $T(x) \varepsilon C - N$ then there is a sequence z in Z such that $T(xz)$ is not in C.*

Consider the transformation T_x defined by $T_x(y) = T(xy)$. Its matrix is (b_{nv}) where $b_{nv} = x_v a_{nv}$. Since $x \varepsilon B$, conditions (i) and (ii) for regularity are satisfied. Since $T(x) \varepsilon C - N$, $\lim T(x) = T_x\text{-lim}\ (1) = c$ exists, and is not zero; the transformation $(1/c)T_x$ is then regular, and there is then a sequence of zeros and ones z for which $T_x(z) = T(xz)$ does not converge.

COROLLARY. *Given T, and a set A in $\mathfrak{D}_T$ with $D_T(A) > 0$, there is a set B contained in A and not in $\mathfrak{D}_T$.*

For, if α is the characteristic function of A, then αz is the characteristic function of a set B contained in A.

We introduce the symbol V for the special transformation which takes the sequence $(x_1, x_2, \cdots)$ into the sequence $(x_2, x_3, \cdots)$.

LEMMA. *Let $\beta \varepsilon Z$, $\alpha \varepsilon Z$, with $\alpha \leq \beta$. Then, if $\beta V(\beta) = 0$ there exists a sequence β^* in Z such that $\alpha = \beta\beta^*$ while $(C,1)\text{-lim}\ (\beta^* - \beta) = 0$.*

Choose β^* as $\alpha + V(\beta) - V(\alpha)$. Since $\alpha \leq \beta$, $\beta\alpha = \alpha$ and $\beta V(\alpha) = \beta V(\beta) V(\alpha) = 0$. Hence, $\beta\beta^* = \beta\alpha + \beta V(\beta) - \beta V(\alpha) = \alpha$. Since $\beta^*(n) = \alpha(n) + \beta(n+1) - \alpha(n+1)$, investigating the possible values of each term, $\beta^*(n)$ is seen to be always either zero or one, and $\beta^* \varepsilon Z$. If $x \varepsilon B$, then $V(x) - x$ is always $(C,1)$ summable to zero; since $\beta^* - \beta = \alpha - \beta + V(\beta) - V(\alpha) = \{V(\beta) - \beta\} - \{V(\alpha) - \alpha\}$, we conclude that $(C,1)\text{-lim}\ (\beta^* - \beta) = 0$.

We use this to prove that a regular summability method cannot preserve products, if it is stronger than $(C,1)$.

THEOREM 16. *If $x \varepsilon Z$, $xV(x) = 0$, and $T(x) \varepsilon C - N$, then there is a sequence y in Z with $T(y) \varepsilon C$ but $T(xy) \notin C$.*

By Theorem 15, there is a sequence z in Z such that $T(xz) \notin C$. Applying the lemma, with $\beta = x$, and $\alpha = xz$ we set $y = \beta^*$. Since $\beta\beta^* = xy = \alpha = xz$, $T(xy) \notin C$. Since T is regular over $(C,1)$, $T\text{-lim}\ (y - x) = 0$, so that $T(y) \varepsilon C$.

COROLLARY. *Let $A \varepsilon \mathfrak{D}_T$ with $D_T(A) > 0$; suppose that A does not contain a pair of consecutive integers. Then there is a set B also in $\mathfrak{D}_T$ but such that $A \cap B$ and $A \cup B$ do not belong to $\mathfrak{D}_T$.*

This is a generalization of the counterexample given for ordinary density in the previous section. In fact, we may again take A as $\{3N\}$. It is therefore true that no generalized density of this type can also be a measure density. Let us now assume that T is positive so that if x is positive, $T\text{-lim inf}\ x \geq 0$.

THEOREM 17. *The Carathéodory closure of $\mathfrak{D}_T$ is again itself.*

Suppose that S belongs to the closure of $\mathfrak{D}_T$. Then, for any positive ε, we can find sets A and B in $\mathfrak{D}_T$ such that $A \subset S \subset B$ and $D_T(B) - D_T(A) < \varepsilon$. Since $\alpha \leq \sigma \leq \beta$, $D_T(A) \leq T\text{-lim inf } \sigma \leq T\text{-lim sup } \sigma \leq D_T(B)$ and therefore $UD_T(S) - LD_T(S) < \epsilon$; letting ϵ approach zero, the upper and lower T densities of S must be equal, and S belongs to $\mathfrak{D}_T$.

If we relax the restriction (iv) that T be stronger than $(C, 1)$, we obtain a generalized density, now, however, not an extension of ordinary density. Corresponding questions may be discussed. Thus, in Theorem 16, we can replace the condition that T be stronger than $(C, 1)$ by the condition

$$\lim_{n\to\infty} \sum_{v=1}^{\infty} | a_{n,v+1} - a_{n,v} | = 0$$

since this is necessary and sufficient that $T\text{-lim} \{V(x) - x\} = 0$ for all x in B. It would also be of interest to know if Theorems 9 and 13 carry over to these generalized densities.

8. The dyadic mapping. Any total measure on the class of subsets of I defines a mapping of this class into the interval $[0, 1]$. Conversely, a measure is such a mapping, which in addition is additive for disjoint sets. Let us consider the function Γ given by $\Gamma(A) = \sum_{1}^{\infty} \alpha(n) 2^{-n}$, where $\alpha(n)$ is the characteristic function of A. This maps the set A onto the real number whose dyadic expansion is given by the sequence of zeros and ones α. It is evident that $\Gamma(A \cup B) + \Gamma(A \cap B) = \Gamma(A) + \Gamma(B)$ and that if $A \subset B$, $\Gamma(A) \leq \Gamma(B)$. Γ is a total measure defined on I; it differs from the ones described in the previous section in that finite sets do not have zero measure. This is also true of any measure given by a linear functional of the form $L(\alpha) = \Sigma \alpha(n) c_n$. Suppose that $\Gamma(A) = \Gamma(B)$. In terms of α and β this means that if $A \neq B$, there must be an n_0 such that $\alpha(n) = \beta(n)$ for $n < n_0$, $\alpha(n_0) = 1$, $\beta(n_0) = 0$, and $\alpha(n) = 0$, $\beta(n) = 1$ for $n > n_0$. Returning to the sets, this implies that $A \doteq 0$, $B \doteq I$. Thus, if we consider only infinite subsets of I, Γ is a $1:1$ mapping of this class onto $(0, 1)$ and at most $2:1$ for the class of all subsets.

We can define a measure on classes of sets of integers by the usual Lebesgue measure of the corresponding sets of points on the unit interval into which these classes are mapped by Γ. For example, consider the class $\mathfrak{D}_0$. Each set A of this is expressible as the union of a finite number of arithmetic progressions with the possible exception of a finite number of points. The

function $\alpha(n)$ is periodic for large n, and $\Gamma(A)$ is rational. Conversely, each rational point of $[0,1]$ is given by a set of $\mathfrak{D}_0$. Thus, $\Gamma(\mathfrak{D}_0)$ is precisely the set of rational points of $[0,1]$. (This affords us another simple proof of the fact that $\mathfrak{D}_0$ is countable.)

In a similar fashion we may investigate $\Gamma(\mathfrak{D}_\mu)$. Since $\mathfrak{D}_\mu$ contains infinite sets of measure zero, and any subset of such a set is also measurable, $\mathfrak{D}_\mu$ has cardinal c as we have observed before. In Section **5**, we proved that a quasi-progression with irrational difference was not measurable μ. If S is a subset of such a set A, then either S or $A - S$ must also be non-measurable, and there are c non-measurable sets. Thus, $\Gamma(\mathfrak{D}_\mu)$ is a non-countable set on the unit interval containing the rational points whose complement is also non-countable. It would be of interest to know if this set is measurable, and if so, whether its measure is zero or one.

The set $\Gamma(\mathfrak{D})$ is more easily discussed; since A belongs to $\mathfrak{D}$ only when the sequence α is $(C,1)$ summable, classical results show that $\Gamma(\mathfrak{D})$ has measure 1, and is of first category in the unit interval [2]. Since $\mathfrak{D}_\mu \subset \mathfrak{D}$, $\Gamma(\mathfrak{D}_\mu)$ too is of first category. Moreover, in this sense, almost every set of integers has density 1/2. Similarly, since $\mathfrak{D} \subset \mathfrak{D}_T$, $\Gamma(\mathfrak{D}_T)$ also has measure 1; however, this set too may be shown to be of first category [5].

SOCIETY OF FELLOWS,
HARVARD UNIVERSITY.

BIBLIOGRAPHY.

1. S. Banach, *Théorie des opérations linéaires.*
2. E. Borel, "Les probabilités dénombrables et leurs applications arithmétiques," Rendiconti del Circolo Mathematico di Palermo vol. 27 (1909), pp. 247-271.
3. Hardy and Wright, *Introduction to the theory of numbers.*
4. T. H. Hildebrandt, "Bounded linear functional operators," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 868-875.
5. J. D. Hill, "Summability of sequences of 0's and 1's," *Annals of Mathematics*, vol. 46 (1945), pp. 556-562.
6. Kestelman, *Modern theory of integration.* (Chapters II and III)
7. G. Polya, "Lücken und Singularitäten von Potenzreihen," *Mathematische Zeitschrift*, vol. 29 (1929), pp. 549-640.
8. H. Steinhaus, "Some remarks on the generalization of limit," (Polish) *Prace Mat.-fiz.*, vol. 22 (1911), pp. 121-134.
9. O. Toeplitz, "Über allgemeine lineare Mittelbildungen," *Prace Mat.-fiz.*, vol. 22 (1911), pp. 113-119.
10. E. B. Van Vleck, "On non-measurable sets with an example," *Transactions of the American Mathematical Society*, vol. 9 (1908), pp. 237-244.
11. A. Wintner, *Measure theory in arithmetical semi-groups.*

AN INEQUALITY FOR POTENTIAL FUNCTIONS.*

By K. O. Friedrichs.

In this paper we shall prove the following inequality: To every domain $\mathfrak{R}$, of a character to be specified, in the N-dimensional space of N variables $x_1, \cdots, x_N$, there is a constant $\Gamma > 0$ such that

$$(1) \qquad \int_{\mathfrak{R}} \phi_1^2 dx \leq \Gamma \int_{\mathfrak{R}} [\phi_2^2 + \cdots + \phi_N^2] dx; \qquad dx = dx_1 \cdots dx_N,$$

holds for the derivatives

$$\phi_k = \partial\phi/\partial x_k$$

of every potential function $\phi(x_1, \cdots, x_N)$ with finite Dirichlet integral,

$$\int_{\mathfrak{R}} \Sigma_\mu \phi_\mu^2 dx < \infty,$$

satisfying the condition

$$(2) \qquad \int_{\mathfrak{R}} \phi_1 dx = 0.$$

Inequality (1) also holds, with an appropriate constant Γ_*, if condition (2) is replaced by

$$(2)_* \qquad \int_{\mathfrak{R}_*} \phi_1 dx = 0,$$

$\mathfrak{R}_*$ being a fixed proper subdomain of $\mathfrak{R}$.

It is clear that the inequality would not hold if condition (2) were omitted, since the potential function $\phi = x_1$ would be a counter-example. If, on the other hand, terms were omitted from the right hand side of (1) the resulting inequality would also not hold: If, for example, ϕ_2^2 were omitted, the function $\phi = x_2^2 - (x_1 - c_1)^2$ would be counter-example, the constant c_1 being so chosen that condition (2) or $(2)_*$ is satisfied.

The inequality (1) is a generalization of the inequality

$$\int\int_{\mathfrak{R}} u^2 dxdy \leq \Gamma \int\int_{\mathfrak{R}} v^2 dxdy$$

* Received March 21, 1946.

for the real and imaginary parts of an analytic function $u + iv$ of the complex variable $x + iy$ under the condition

$$\iint_{\mathfrak{R}} u\,dx\,dy = 0.$$

This inequality was proved earlier by the author[1]; it may be considered an analogue of the inequality

$$\int_{\mathfrak{B}} u^2 ds \leq \Gamma \int_{\mathfrak{B}} v^2 ds$$

if

$$\int_{\mathfrak{B}} u\,ds = 0$$

for the boundary $\mathfrak{B}$ of a circle, proved by Hilbert [2] and Lichtenstein [3], and generalized by M. Riesz [4], [5].

Our inequality is one of a class of inequalities which can be proved by similar methods. We mention *Korn's inequality* which we shall prove in a forthcoming paper.[2] It refers to systems $[u_1, \cdots, u_N]$ of N functions in $\mathfrak{R}$ with finite Dirichlet integral and involves the symmetric and the antisymmetric part of the gradient tensor $u_{k/l} = \partial u_k/\partial x_l$:

$$s_{kl} = \tfrac{1}{2}(u_{k/l} + u_{l/k}), \qquad r_{kl} = \tfrac{1}{2}(u_{k/l} - u_{l/k}).$$

The statement is: To every domain $\mathfrak{R}$, of a character to be specified, there is a constant $\mathrm{K} > 0$ such that

$$\int_{\mathfrak{R}} \Sigma_{\kappa,\lambda} r_{\kappa\lambda}{}^2 dx \leq \mathrm{K} \int_{\mathfrak{R}} \Sigma_{\kappa,\lambda} s_{\kappa\lambda}{}^2 dx$$

holds for all functions $[u_1, \cdots u_N]$ which satisfy the side condition

$$\int_{\mathfrak{R}} r_{kl} dx = 0, \qquad k, l = 1, \cdots, N.$$

(The functions u_k are not required to satisfy a differential equation.) This inequality, given by Korn in 1909, furnishes an estimate of the Dirichlet integral by its "symmetric part":

$$\int_{\mathfrak{R}} \Sigma_{\kappa,\lambda} u_{\kappa/\lambda}{}^2 dx \leq (\mathrm{K} + 1) \int_{\mathfrak{R}} \Sigma_{\kappa,\lambda} s_{\kappa\lambda}{}^2 dx$$

---

[1] See [1] in the Bibliography. The present proof, when specialized to $N = 2$, is simpler than that given in [1].

[2] On the Boundary Value Problems of the Theory of Elasticity and Korn's Inequality.

and that is the reason why it plays such a decisive part in the theory of elasticity.

Our inequality (1) under the condition (2) and also Korn's inequality will be established for *domains* $\mathcal{R}$ characterized by a particular set of requirements: Firstly, $\mathcal{R}$ is open and connected. Next, there exists a vector $\Omega_k(x_1, \cdots, x_n)$ with the following properties:

1. Ω_k is continuous in $\mathcal{R} + \mathcal{B}$, $\mathcal{B}$ being the boundary of $\mathcal{R}$,

2. $\Omega_k = 0$ on $\mathcal{B}$,

3. Ω_k has continuous first derivatives $\Omega_{k/m} = \partial\Omega_k/\partial x_m$ in $\mathcal{R}$,

4. These derivatives are bounded in $\mathcal{R}$:

$$|\Omega_{k/m}| \leq C_1 \quad \text{in } \mathcal{R}.$$

Denoting by $\mathcal{R}_\delta$ the subdomain of all points in $\mathcal{R}$ whose distance from $\mathcal{B}$ is greater than δ we require:

5. For a certain $v > 0$, for which $\mathcal{R}_v$ is connected, we have

$$\Omega_{\kappa/\kappa} \geq 1 \quad \text{in} \quad \mathcal{R} - \mathcal{R}_v.$$

It is finally required that for every $\delta > 0$, $\delta \leq v$, there exist a function H

$$\mathrm{H} = \mathrm{H}^\delta(x_1, \cdots x_N)$$

with continuous first derivatives H_k such that

1. $\mathrm{H} = 1 \quad \text{in } \mathcal{R}_\delta,$

2. $\mathrm{H} \geq 0 \quad \text{in } \mathcal{R},$

3. $\mathrm{H} = 0 \quad \text{in } \mathcal{R} - \mathcal{R}_{\delta'}$ for an appropriate positive $\delta' < \delta$,

4. $|\mathrm{H}_k| \leq C_2/\delta$, with an appropriate constant C_2 (depending on $\mathcal{R}$).

Domains satisfying these requirements will be called Ω-domains. In the second section of this paper we shall show that a domain whose boundary possesses a continuously differentiable normal-vector is an Ω-domain. We shall also show that domains with corners or edges, such as parallelopipeds, are Ω-domains.

We mention that these domains $\mathcal{R}$ treated in **2** have an additional property, viz. that there exists a set of Ω-domains $\mathcal{R}^{(\delta)}$ for $0 \leq \delta \leq \delta_0$ (with a certain $\delta_0 > 0$) with $\mathcal{R}^{(0)} = \mathcal{R}$, and $\mathcal{R}_\delta \subset \mathcal{R}^{(\delta)} \subset \mathcal{R}$, such that

the derivatives of the corresponding functions $\Omega^{(\delta)}$ are uniformly bounded, (i. e., independent of δ in $0 \leq \delta \leq \delta_0$). Domains enjoying this property will be called special Ω-domains. The proof of inequality (1) presented in **1** will imply that also the constants $\Gamma^{(\delta)}$ in inequality (1) are uniformly bounded for such domains. Then we have as an immediate consequence the

THEOREM. *Let $\mathcal{R}$ be a special Ω-domain. Then for every potential function ϕ with continuous derivatives for which $\int_{\mathcal{R}} [\phi_2{}^2 + \cdots + \phi_N{}^2]dx$ is finite, also $\int_{\mathcal{R}} \phi_1{}^2 dx$ is finite.*

This theorem is of special interest for the case $N = 2$ where it implies that for every analytic function $w = u + iv$ of $z = x + iy$ for which $\int\int_{\mathcal{R}} u^2 dxdy$ is finite for a special Ω-domain $\mathcal{R}$, also $\int\int_{\mathcal{R}} v^2 dxdy$ is finite.[3]

1. Proof of the inequality (1). The proof of our inequality proceeds in four steps. The first decisive step consists in establishing the inequality

$$\int_{\mathcal{R}} \phi_1{}^2 dx \leq \Gamma_1 \int_{\mathcal{R}_v} \phi_1{}^2 dx + \Gamma_1 \int_{\mathcal{R}} \sum_{\nu=2}^{N} \phi_\nu{}^2 dx; \tag{1.01}$$

more precisely, we shall show that to every $\mathcal{R}_v$ there is a constant $\Gamma > 0$ such that (1.01) holds for all potential functions with finite Dirichlet integral. The later steps are concerned with establishing the inequality

$$\int_{\mathcal{R}_v} \phi_1{}^2 dx \leq \Gamma_2 \int_{\mathcal{R}} \sum_{\nu=2}^{N} \phi_\nu{}^2 dx \tag{1.02}$$

under condition (2).

The proof of the inequality (1.01) is derived from *the identity*

$$\begin{aligned} O^m &\equiv (\phi^2{}_m - \sum_{\kappa \neq m} \phi_\kappa{}^2)_m + 2 \sum_{\kappa \neq m} (\phi_\kappa \phi_m)_\kappa \\ &\equiv 2\, \Sigma_\kappa (\phi_\kappa \phi_m)_\kappa - (\Sigma_\kappa\, \phi_\kappa{}^2)_m \\ &\equiv 2\phi_m \Sigma_\kappa\, \phi_{\kappa\kappa} = 0 \end{aligned} \tag{1.03}$$

valid for potential functions $\phi(x_1, \cdots x_N)$. Subscripts $m, \cdots$ here and

[3] This statement together with (1) entails: If a potential function u can be approximated in the mean by potential functions with continuous derivatives in $\mathcal{R} + \mathcal{B}$, then the same is true for the conjugate function. For this statement under conditions for the boundary somewhat different from ours in particular also for star-domains, see N. Aronszajn [6].

throughout indicate differentiation with respect to x_m; (except for the components Ω, for which the subscripts indicating differentiation are separated by a stroke).

With the abbreviation

$$\Sigma'_{\mu} = \sum_{\mu \neq 1}, \qquad \Sigma''_{\mu,\nu} = \sum_{\mu\neq 1, \nu \neq 1} \tag{1.04}$$

we consider the relation

$$\int_{\mathcal{R}} \{\Omega_1 O^1 - \Sigma'_{\mu}\Omega_{\mu} O^{\mu}\} dx = 0, \tag{1.05}$$

which, if integration by parts could be carried out, would go over into

$$\int_{\mathcal{R}} \{\Omega_{1/1}(\phi_1{}^2 - \Sigma'_{\nu}\phi_{\nu}{}^2) + 2\,\Sigma'_{\nu}\Omega_{1/\nu}\phi_1\phi_{\nu} + \Sigma'_{\mu}\Omega_{\mu/\mu}(\phi_1{}^2 + \Sigma'_{\nu}\phi_{\nu}{}^2)$$
$$- 2\,\Sigma'_{\mu}\Omega_{\mu/1}\phi_{\mu}\phi_1 - 2\,\Sigma''\Omega_{\mu/\nu}\phi_{\mu}\phi_{\nu}\} dx = 0,$$

or

$$\int_{\mathcal{R}} \Sigma_{\kappa}\Omega_{\kappa/\kappa}\phi_1{}^2 dx - \int_{\mathcal{R}} [2\,\Sigma'_{\mu}(\Omega_{\mu/1} - \Omega_{1/\mu})\phi_{\mu}\phi_1 + (\Omega_{1/1} - \Sigma'_{\mu}\Omega_{\mu/\mu})\Sigma'_{\nu}\phi_{\nu}{}^2 + 2\,\Sigma''_{\mu,\nu}\Omega_{\mu/\nu}\phi_{\mu}\phi_{\nu}] dx = 0. \tag{1.06}$$

That integration by parts is valid is not obvious, however, since no asumptions were made about the behavior of the derivatives ϕ_k at the boundary $\mathcal{B}$, except that the Dirichlet integral be finite. Instead of relation (1.05) we, therefore, consider first the relation

$$\int_{\mathcal{R}} \mathrm{H}[\Omega_1 O^1 - \Sigma'_{\mu}\Omega_{\mu} O^{\mu}] dx = 0, \tag{1.05$'$}$$

in which $\mathrm{H} = \mathrm{H}^{\delta}$ is the function postulated in the introduction. Since $\mathrm{H}^{\delta} = 0$ in $\mathcal{R} - \mathcal{R}_{\delta}$, we may integrate by parts and obtain a relation which differs from (1.06) only in that $\Omega_{k/m}$ is replaced by $(\mathrm{H}^{\delta}\Omega_k)_m$. These derivatives are bounded in $\mathcal{R} - \mathcal{R}_{\delta}$. For $\Omega_{k/m}$ is assumed to be bounded; further, since $\Omega_k = 0$ on $\mathcal{B}$, $\delta^{-1}\Omega_k$ is bounded in $\mathcal{R} - \mathcal{R}_{\delta}$, hence also $\mathrm{H}^{\delta}{}_m\Omega_k = (\delta \mathrm{H}^{\delta}{}_m)(\delta^{-1}\Omega_k)$ by assumption 4 on H^{δ}. Since $\mathrm{H}^{\delta} = 1$ in $\mathcal{R}_{\delta}$ we obtain for the left member of (1.06), extended over $\mathcal{R}_{\delta}$ instead of $\mathcal{R}$, the estimate

$$\Gamma_3 \int_{\mathcal{R} - \mathcal{R}_{\delta}} \Sigma_{\mu}\phi_{\mu}{}^2 dx,$$

which approaches zero with δ. Therefore relation (1.06) is valid.

From relation (1.06) and properties 4 and 5 of Ω_k we obtain, after using Schwarz's inequality,

$$\int_{\mathfrak{R}-\mathfrak{R}_v} \phi_1^2 dx \leq \Gamma_4 \int_{\mathfrak{R}_v} \phi_1^2 dx + \Gamma_4 \int_{\mathfrak{R}} \Sigma'_\nu \phi_\nu^2 dx$$
$$+ \Gamma_3 \{ \int_{\mathfrak{R}} \phi_1^2 dx \int_{\mathfrak{R}} \Sigma'_\nu \phi_\nu^2 dx \}^{\frac{1}{2}}$$

whence, after a few obvious steps,

$$\int_{\mathfrak{R}} \phi_1^2 dx \leq 2(\Gamma_4 + 1) \int_{\mathfrak{R}_v} \phi_1^2 dx + (2\Gamma_4 + \Gamma_3^2) \int_{\mathfrak{R}} \Sigma'_\nu \phi_\nu^2 dx. \tag{1.07}$$

which is equivalent to (1.10).

To prove our theorem we therefore need only estimate $\int_{\mathfrak{R}_v} \phi_1^2 dx$ in terms of $\int_{\mathfrak{R}} \Sigma'_\nu \phi_\nu^2 dx$. This could be done in several ways. Rather directly we proceed in three steps as follows:

First we employ *Poincaré's inequality* in a restricted sense:[4] There is a constant C_* such that

$$\int_{\mathfrak{R}_v} \psi^2 dx \leq C_* \int_{\mathfrak{R}_v} \Sigma_\mu \psi_\mu^2 dx$$

holds for all functions ψ possessing a finite Dirichlet integal over $\mathfrak{R}$ and satisfying the relation

$$\int_{\mathfrak{R}_*} \psi dx = 0,$$

where $\mathfrak{R}_*$ is a subdomain of $\mathfrak{R}_v$. We apply this inequality to $\psi = \phi_1$ and obtain

$$\int_{\mathfrak{R}_v} \phi_1^2 dx \leq C_* \int_{\mathfrak{R}_v} \Sigma_\mu \psi_{1\mu}^2 dx \tag{1.08}$$

$$\int_{\mathfrak{R}_*} \phi_1 dx = 0. \tag{1.09}$$

Poincaré's inequality estimates the integral of the square of the function in terms of the integral of the sum of the squares of the derivatives. For potential functions estimates in the opposite direction are possible: the integral of the sum of the squares of the derivatives can be estimated through the integral of the square of the function over a wider domain. More specifically, we need an estimate of the integral

$$\int_{\mathfrak{R}_v} \Sigma_\mu \phi_{1\mu}^2 dx \quad \text{through} \quad \int_{\mathfrak{R}_v} \Sigma'_\nu \phi_\nu^2 dx.$$

---

[4] See [6] Courant-Hilbert, vol. II, ch. VII, §§ 3.1, 6, 3.8. At this place the assumption is used that $\mathfrak{R}_v$ is connected.

To this end we employ the identity

$$(1.10)\qquad \begin{aligned}&(\Sigma'_\nu\phi_\nu{}^2)_{11}-\Sigma'_\mu(\Sigma'_\nu\phi_\nu{}^2)_{\mu\mu}+2\,\Sigma''_{\mu,\nu}(\phi_\mu\phi_\nu)_{\mu\nu}\\&=2[\Sigma'_\nu\phi_{\nu 1}{}^2-\Sigma''_{\mu,\nu}\phi_{\mu\mu}\phi_{\nu\nu}]+2[\Sigma'_\nu\phi_\nu\phi_{\nu 11}+\Sigma''_{\mu,\nu}\phi_\nu\phi_{\mu\mu\nu}]\\&=2\,\Sigma_\kappa\phi_{\kappa 1}{}^2,\end{aligned}$$

which is valid for potential functions ϕ. Let $H=H^\nu$ be the function postulated in the introduction which vanishes in $\mathcal{R}-\mathcal{R}_\nu$. Then we multiply (1.10) by H, integrate over $\mathcal{R}$, and apply integration by parts. Thus we obtain

$$(1.11)\qquad 2\int_{\mathcal{R}} H\Sigma_\kappa\phi_{\kappa 1}{}^2dx=\int_{\mathcal{R}}[(H_{11}-\Sigma'_\mu H_{\mu\mu})\Sigma'_\nu\phi_\nu{}^2+2\,\Sigma''_{\mu,\nu}H_{\mu\nu}\phi_\mu\phi_\nu]\,dx.$$

Since $H=1$ in $\mathcal{R}_\nu$, $H\geqq 0$ in $\mathcal{R}$, and the second derivative H_{mn} are bounded we can derive from (1.11) the relation

$$(1.12)\qquad \int_{\mathcal{R}_\nu}\Sigma_\kappa\phi_{\kappa 1}{}^2dx\leqq\Gamma_5\int_{\mathcal{R}_\nu}\Sigma'_\nu\phi_\nu{}^2dx.$$

Combining (1.12) with (1.08) and (1.01) we find

$$(1.13)\qquad \int_{\mathcal{R}}\phi_1{}^2dx\leqq\Gamma_6\int_{\mathcal{R}}\Sigma'\phi_\nu{}^2dx,$$

if (1.09) or $(2)_*$ for a proper subdomain $\mathcal{R}_*$ holds. If the side condition (2) holds for the whole domain $\mathcal{R}$ we assign to the function ϕ a constant c such that $\phi^*=\phi+cx$ satisfies (1.09). Inequality (1.13) then holds for ϕ^*. Since by (2)

$$\int_{\mathcal{R}}\phi_1{}^2dx=\int_{\mathcal{R}}(\phi^*{}_1)^2dx-c^2\int_{\mathcal{R}}x^2dx\leqq\int_{\mathcal{R}}(\phi^*{}_1)^2dx$$

and the right hand member in (1.13) is the same for ϕ as for ϕ^*, we see that (1.13) also holds for the function ϕ satisfying relation (2). Thus our inequality (1) is proved.

2. On the Admitted Domains. In this section we shall discuss simple conditions for domains $\mathcal{R}$ to be what we have called Ω-domains in the introduction.

We first state: The domain $\mathcal{R}$ is an Ω-domain if there exists an inside neighborhood $\mathcal{S}^0$ of its boundary $\mathcal{B}$ which is a one-to-one image of a spherical shell $\rho_0 \leq \rho \leq 1$, $\rho^2 = \Sigma_\kappa \xi_\kappa^2$, of a $[\xi_1, \cdots \xi_N]$-space given by functions $x = x(\xi)$ which possess continuous first derivatives $x_{l/m} = \partial x_l / \partial \xi_m$ in $\rho_0 \leq \rho \leq 1$ such that the jacobian does not vanish:

$$\det x_{l/m} \neq 0 \quad \text{in} \quad \rho_0 \leqq \rho \leqq 1. \tag{2.01}$$

The boundary $\mathcal{B}$ is supposed to be the image of $\rho = 1$.

To prove this statement we have to construct functions Ω_k and H^δ, satisfying the specifications set up in the introduction.

We first observe that the mapping has an inverse $\xi_k = \xi_k(x)$ with continuous derivatives $\xi_{k/l} = \partial \xi_k / \partial x_l$ in $\mathcal{S}^0$. We have

$$\Sigma_\lambda \xi_{k/\lambda} x_{\lambda/m} = \delta_{km}. \tag{2.02}$$

If the function $x_l(\xi)$ possessed continuous second derivatives $x_{l/mn}$ we could simply set

$$\Omega_l = \Sigma_\mu \xi_\mu (1 - \rho^{-N}) x_{l/\mu},$$

because then

$$\Sigma_\lambda \Omega_{\lambda/\lambda} = N + (1 - \rho^{-N}) \Sigma_{\lambda,\mu,\nu} \xi_\mu \xi_{\nu/\lambda} x_{\lambda/\mu\nu}$$

would be greater than 1 in the neighborhood of $\rho = 1$, ($N \geq 2$ being assumed).

Without assuming the existence of derivatives $x_{l/mn}$ we proceed as follows. Clearly, we can approximate the functions $x_l(\xi)$ by functions $X_l(\xi)$ defined in a shell $\rho_1 \leq \rho \leq 1$, $\rho_0 < \rho_1 < 1$, which possesses continuous second derivatives $X_{l/mn}$ and whose first derivatives $X_{l/mn}$ approximate $x_{l/m}$ as closely as desired. For our purpose it is sufficient to require

$$\begin{aligned} &| \Sigma_{\kappa,\lambda,\mu} \xi_\mu \xi_\kappa \xi_{\kappa/\lambda} (X_{\lambda/\mu} - x_{\lambda/\mu}) | \\ &\qquad < (2N)^{-1} \rho_1^{N+2} \quad \text{in} \quad \rho_1 \leq \rho \leq 1. \end{aligned} \tag{2.03}$$

Now let Z be a function of ξ with continuous derivatives which equals zero for $\rho \leq \rho_1$ and equals 1 in the shell $\rho_2 \leq \rho \leq 1$, with $\rho_1 < \rho_2 < 1$. We may consider Z a function of x in $\mathcal{R}$ with

$$\mathrm{Z} = 0 \quad \text{in} \quad \mathcal{R} - \mathcal{S}^1, \ \mathrm{Z} = 1 \quad \text{in} \quad \mathcal{S}^2, \tag{2.04}$$

$\mathcal{S}^1$ and $\mathcal{S}^2$ being the images of the shells $\rho_1 < \rho < 1$ and $\rho_2 \leq \rho \leq 1$. Then we set

$$\Omega_l = 2\, Z\Sigma_\mu \xi_\mu (1 - \rho^{-N}) X_{l/\mu}. \tag{2.05}$$

Clearly, these functions Ω_l have bounded continuous derivatives in $\mathcal{R}$ and vanish on $\mathcal{B}$. In $\mathcal{S}^2$ we have in particular

$$\begin{aligned} \Sigma_\lambda \Omega_{\lambda/\lambda} = {} & 2N\Sigma_{\kappa,\lambda,\mu} \xi_\mu \xi_\kappa \rho^{-N-2} \xi_{\kappa/\mu} (X_{\lambda/\mu} - x_{\lambda/\mu}) \\ & + (1 - \rho^{-N}) B, \end{aligned} \tag{2.06}$$

where B is a bounded function. By virtue of assumption (2.03) relation (2.06) yields

$$\Sigma_\lambda \Omega_{\lambda/\lambda} \geq N - (1 - \rho_1^{-N}) \mid B \mid.$$

Hence a value ρ_3 with $\rho_2 < \rho_3 < 1$ can be found such that $\Sigma_\lambda \Omega_{\lambda/\lambda} \geq 1$ in $\mathcal{S}^3$, the image of $\rho_3 \leq \rho \leq 1$. Clearly a value of v can be found such that $\mathcal{S}^3$ contains $\mathcal{R} - \mathcal{R}_v$.

To construct the function H^δ described in the introduction we introduce the shells $\mathcal{S}_\epsilon$ adjacent to the boundary $\mathcal{B}$ as images of the shells $1 - \epsilon \leq \rho \leq 1$ in the ξ-space. With

$$t^2 = \max.\ \Sigma_{\lambda,\mu} x^2 x^2{}_{\lambda/\mu}, \text{ in } \rho_0 \leq \rho \leq 1, \tag{2.07}$$

and

$$\tau^{-2} = \max.\ \Sigma_{\kappa,\lambda} \xi^2{}_{\kappa/\lambda}, \text{ in } \mathcal{S}_{1-\rho_0}$$

we state that

$$\mathcal{R} - \mathcal{R}_{\tau\epsilon} < \mathcal{S}_\epsilon < \mathcal{R} - \mathcal{R}_{\tau\epsilon}.$$

Let P and P_ϵ be the images of two points on $\rho = 1$ and $\rho = 1 - \epsilon$ on the same ray in the ξ-space. Then, by (2.07), clearly $\mid P - P_\epsilon \mid \leq t\epsilon$ which implies that $\mathcal{S}$ lies in $\mathcal{R} - \mathcal{R}_{\tau\epsilon}$. Similarly one finds $\mathcal{S}_\epsilon > \mathcal{R} - \mathcal{R}_{\tau\epsilon}$ from (2.08).

To a given $\delta > 0$ we now set $\delta' = \delta\tau/2t$. Then we have

$$\mathcal{R} - \mathcal{R}_{\delta'} < \mathcal{S}_{\delta'/\tau} < \mathcal{S}_{\delta/t} < \mathcal{R} - \mathcal{R}_\delta.$$

We then choose H as a function of ξ with continuous derivatives such that $H = 1$ for $1 - \rho \geq \rho/t$ while $H = 0$ for $\delta'/\tau = \delta/2t > 1 - \rho \geq 0$. We can do this in such a manner that the derivatives of H with respect to $[\xi]$ are in absolute value less than $3t/\delta$. When we now consider H a function of x it is

clear that $\mathrm{H} = 0$ in $\mathcal{S}_{\delta'/\tau}$, hence in $\mathcal{R} - \mathcal{R}_\delta$, $\mathrm{H} = 1$ in $\mathcal{R} - \mathcal{S}_{\delta/t}$, hence in $\mathcal{R}_\delta$, and that the derivatives of H with respect to x_1 are less than $3N\,ht/\delta$ when $h = \max |\xi_{k/l}|$ in $\mathcal{S}^0$. Thus the function $\mathrm{H} = \mathrm{H}^\delta$ satisfies the specifications set up in the introduction.

We note that a domain is also an Ω-domain if its boundary consists of a finite number of pieces each of the character described at the beginning of this section.

It is evident that the domains here considered are also special Ω-domains (see the end of the introduction); we need only set $\mathcal{R}^{(\delta)} = \mathcal{R} - \mathcal{S}_{\delta/t}$. For these domains the same approximating functions $X(\xi)$ can be used: $X^{(\delta)}(\xi) = X((1-\epsilon)\xi)$. It is then obvious that the derivatives of the functions $\Omega^{(\delta)}$ are uniformly bounded.

We proceed to show that domains *with corners or edges* are also Ω-domains. We first consider the case that the domain $\mathcal{R}$ has a corner or an edge at the origin O in such a way that, in a neighborhood of O, $\mathcal{R}$ can be represented by

$$0 < x_1, 0 < x_2 \cdots, 0 < x_n; \text{ for } n \leq N.$$

We then must construct a vector Ω_k defined for $0 \leq x_1, \cdots, x_n$, which is continuous up to the boundary $x_1 x_2 \cdots x_n = 0$ and vanishes there, which in $0 \leq x_1, \cdots, x_n$ possesses continuous bounded derivatives, and for which

$$\Omega_{1/1} + \cdots + \Omega_{n/n} \geq 1. \tag{2.09}$$

Such a vector is

$$\begin{aligned} \Omega_1 = \cdots = \Omega_n &= n[x_1^{-1} + \cdots + x_n^{-1}]^{-1}, \\ &\qquad \text{for } 0 < x_1, \cdots, x_n; \\ &= 0, \\ &\qquad \text{for } x_1 x_2 \cdots x_n = 0. \end{aligned} \tag{2.10}$$

In other words all components are equal and just the harmonic mean of the coordinates. To verify that this vector has the desired properties we first note the relation

$$|\Omega_k| \leq n \min x_\nu, \qquad 0 < x_1, \cdots, x_n;$$
$$\nu = 1, \cdots, n;$$

which is immediately derived from the definition (2.10) and implies that $\Omega_k \to 0$ uniformly as x approaches the boundary $x_1 \cdots x_n = 0$. We further find

$$\Omega_{k/l} = n x_l^{-2}[x_1^{-1} \cdots + x_n^{-1}]^{-2}, \qquad l = 1, \cdots, n,$$
$$= 0 \qquad\qquad l > n,$$

whence $$0 \leq \Omega_{k/l} \leq n \text{ in } 0 < x_1, \cdots, x_n.$$

Finally, (2. 09) obtains,

$$\Omega_{1/1} + \cdots + \Omega_{n/n} = n[x_1^{-2} + \cdots + x_n^{-2}][x_1^{-1} + \cdots + x_n^{-1}]^{-2} \geq 1.$$

The existence of function H^δ with the desired properties is easily seen. Introducing a function $\eta(\xi)$ of a single variable ξ with continuous second derivatives for $0 < \xi$ and the properties $0 \leq \eta(\xi) \leq 1$

$$\eta(\xi) = 0 \quad \text{for } \xi \leq 1/2$$
$$= 1 \quad \text{for } \xi \geq 1,$$

we set

$$H^\delta = \eta(x_1/\delta) \cdots \eta(x_n/\delta).$$

This function has evidently the desired properties with $\delta' = \delta/2$.

If the domain $\mathcal{R}$ is a cube

$$0 < x_1, \cdots, x_N \leq 1$$

we define the vector $\Omega_k^{(m)}$ with $n = N$ for each of the 2^N corners $\mathcal{C}^{(m)}$ $m = 1, \cdots, 2^N$. For the corner $\mathcal{C}^{(1)} = 0$ we define $\zeta(x)$ as a function with continuous derivatives such that

$$0 \leq \zeta(\xi) \leq 1, \quad 0 \leq \xi \leq 1,$$
$$\zeta(\xi) = 1, \quad 0 \leq \xi \leq \tfrac{2}{3},$$
$$= 0, \quad \tfrac{5}{6} \leq \xi \leq 1.$$

We introduce the function $Z^{(1)}(x)$ by

$$Z^{(1)}(x) = \zeta(x_1) \cdots \zeta(x_N).$$

and define the functions $Z^{(2)}(x), \cdots, Z^{(2^N)}(x)$ correspondingly. Then we set

$$\Omega_1 = \cdots = \Omega_N = Z^{(1)}\Omega^{(1)} + \cdots + Z^{(2^N)}\Omega^{(2^N)}.$$

One immediately verifies that this vector has the desired properties. The

functions H^δ for the cube are constructed from those for a single corner in a similar manner. Thus it is seen that *a cube is an* Ω-domain. Evidently a cube is also a special Ω-domain.

Suppose that the corner in *not rectangular* as assumed so far but that it can be obtained from a rectangular corner by a transformation $x_l = x_l(x^*)$ with positive jacobian and bounded second derivatives. Then one sets, in obvious notation

$$\Omega_k = J_0 J^{-1} \Sigma_\lambda \Omega^*{}_\lambda \partial x_k / \partial x_\lambda$$

where J_0 is the minimum of J in the neighborhood considered. By a well known invariance property of the divergence

$$\Sigma_\kappa \partial \Omega_\kappa / \partial x_\kappa = J_0 J^{-1} \Sigma_\lambda \partial \Omega^*{}_\lambda / \partial x^*{}_\lambda \geq J_0 J^{-1} \leq 1,$$

and hence Ω_k enjoys the desired properties. The construction of the functions H^δ is obvious. Thus it is shown that also certain domains with non-rectangular corners or edges are Ω-domains, for which our inequality holds.

NEW YORK UNIVERSITY.

TOPOLOGIES FOR HOMEOMORPHISM GROUPS.*

By Richard Arens.

1. Introduction. This paper investigates the topological space obtained by defining, in a group H of homeomorphisms [1] of a locally compact space A, one or the other of the following two topologies:

a) The k-topology for H is based on neighborhoods U of the following description: Let K and W be compact and open subsets, respectively, of A, and let (K, W) denote the collection of all elements of H which transform K onto a part of W. Let U be the intersection in H of any finite number of all possible such sets, (K, W).

b) The g-topology for H is based on neighborhoods U of the following nature: Let K and W be closed and open subsets, respectively, of A, with the condition that either K, or the complement of W in A, is compact. Let (K, W) contain all those elements of H which send K into W, and let U be the intersection of an arbitrary finite set of such sets as (K, W).

With the k-topology, the binary group operation [2] in H becomes continuous. With the g-topology, we have moreover a continuous group operation of inversion whence H forms a topological group [3] with the g-topology.

In fact, the g-topology is the strongest [4] topology which makes H into a topological group and at the same time insures that $h(x)$ depends continuously on x and h simultaneously.

The k- and g-topologies coincide if A is compact. If A is also metric, these topologies coincide with that studied, for example, by G. Birkhoff [3, p. 872] for the group H of homeomorphisms of a compact metric space,

* Received January 24, 1946. Presented to the American Mathematical Society, December 28, 1942.

[1] For this, and other general topological terms, we follow the usage of Lefschetz (reference [8] in the list at the end of this paper).

[2] In which, for f, g, h, in H, one has $f = gh$ if and only if, for each x in A, $f(x) = g(h(x))$.

[3] Which is not usually complete (in the sense of Weil, [15, p. 31]) but satisfies a condition (see 7) which insures that H is of the second category [2, p. 13] at least if A is perfectly separable.

[4] I. e., providing the most limit elements.

where one metrizes H by setting (f, g) equal to the maximum of the distance from $f(x)$ to $g(x)$, for all x in A.

van der Waerden and van Dantzig [14, p. 370] apply this method to the locally compact, separable metric case by first adding a point to A to make it compact, and metrizing as above. This produces the g-topology. We show that, in general, the g-topologized group H of homeomorphisms of A is topologically isomorphic to the k-topologized group H^* of homeomorphisms of A^* leaving I fixed, where $A^* = A + I$ is the compactification [8, p. 23] of A.

Another result is that if h_ν is a directed set of homeomorphisms of a locally compact, *locally connected* space, which converges to the identity in the k-topology, then the directed set h_ν^{-1} also converges to the identity, in the k-topology. Thus, for homeomorphism groups of locally compact, locally connected spaces, the g- and k-topologies are equivalent.

If H is a compact and effective transformation group (Cf. [9, p. 365]) its topology must coincide with the k- and g-topologies; this is true in particular for compact Lie groups of transformations. For locally compact Lie groups of transformations, the maximal connected subsets of neighborhoods in the g-topology provide a basis for the usual parameter-space topology.

Directed sets of functions, f_ν, defined on a locally compact Hausdorff space, have the following convergence properties involving the k- and g-topologies: [5]

i) f_ν converges to f in the k-topology if and only if $f_\nu(x_\delta)$ converges to $f(x)$ whenever the directed set x_δ converges to x in the domain.

ii) f_ν converges to f in the g-topology if and only if f_ν converges to f, and f_ν^{-1} converges to f^{-1}, in the k-topology, for homeomorphisms.

iii) f_ν converges to f in the g-topology if and only if the convergence of x_δ to x implies and follows from the convergence of $f_\nu(x_\delta)$ to $f(x)$.

We generalize a result of van Dantzig and van der Waerden in showing that an equicontinuous group [6] of homeomorphisms of a connected locally compact space is a locally compact topological group.

---

[5] Cf. a separate and more complete investigation of the k-topology in Arens [1]. See also [14].

[6] I. e., a group whose elements as a whole form an equicontinuous family in the usual metric sense, or in a more general sense expressible by means of uniform structure (Cf. 7 below). For the way in which compactness properties of the group imply equicontinuity, see Eilenberg [6, p. 77].

We formulate a condition upon a topological group of homeomorphisms which insures that the quotient space H/F, where F is the subgroup of elements leaving a point x of A fixed, should be homeomorphic to the domain of transitivity (under H) which contains x: If U be any open set in H, then the set of all $h(x)$, where $h \epsilon U$, is open relative to its domain of transitivity in A. This condition is satisfied if H is compact;[7] and we prove it true also for any transitive Lie group of transformations.[8]

2. Definition and formal interrelation of k- and g-topologies. Let A be any topological space, and let C be any class of continuous functions defined on A and with values also in A. If K and W are subsets of A, we shall always use (K, W) to denote the collection of functions $f \epsilon C$ for which $f(K) \subset W$. If $K_1, W_1, \cdots, K_n, W_n$, where n is finite, are n pairs of subsets of A, we shall denote by $(K_1, \cdots, K_n; W_1, \cdots, W_n)$ the set of all functions $f \epsilon C$ for which $f(K_i) \subset W_i$, $i = 1, 2, \cdots, n$.

DEFINITION. The topology obtained in C by taking the collection of all the sets $(K_1, \cdots, K_n; W_1, \cdots, W_n)$ as neighborhoods of any functions they contain,[9] where the K's are compact and the W's are open, will be called the k-topology for C.[10]

DEFINITION. The topology obtained in C by taking the collection of all the sets $(K_1, \cdots, K_n; W_1, \cdots, W_n)$ as neighborhoods of any functions they contain,[9] where the K's are closed and the W's are open, and, for each i, either K_i or the complement W'_i of W_i is compact, will be called the g-topology for C.

We may adjoin a point I to A, giving $A^* = A + I$, which is compact, by defining the neighborhoods of I to be the exteriors of compact sets of A [8, p. 23]. We may then regard any function $f \epsilon C$ as corresponding to a function f^* defined on A^*, but mapping I, and only I, on I. We thus arrive at a class C^* of continuous functions each mapping A^* into itself, which is in obvious $1 - 1$ correspondence with C.

---

[7] Cf. Montgomery and Samelson [10, p. 455].

[8] We give an example of a compact subset of 3-space for whose group of homeomorphisms H the condition fails for every topology for H.

[9] Or, equivalently, taking the sets (K, W) as a subbase [8, p. 6]. In proofs, whenever it involves no loss in generality, we shall tacitly limit ourselves to the consideration of subbasic neighborhoods.

[10] This topology has been defined, in one way or another, for various special cases by many authors, such as Banach [2, p. 11], G. Birkhoff, *loc. cit.*, Fox [7, p. 429], Pontrjagin [12, p. 127].

Lemma 1. *C with the g-topology, and C^* with the k-topology are homeomorphic under the obvious correspondence.*[11]

The proof of this is rather obvious, and will be omitted.

From this we can infer at once:

Theorem 1. *The g-topologized group H of homeomorphisms of a locally compact Hausdorff space A is topologically isomorphic to the group H^* of those homeomorphisms of the compactification A^* of A which leave the ideal point fixed.*[12]

Of course, it is evident from the definition that the k- and g-topologies coincide when A is compact.

3. A minimal property of the k-topology.

Definition. A topology for C (cf. **2**) will be called *admissible,* if with $x \in A$ and $f \in C$, $F(f, x) = f(x)$ is a continuous function on $C \times A$ into A.

This is equivalent to saying that whenever $f \in C$ and $x \in A$, and a neighborhood W of $f(x)$ are given, we can find a neighborhood $V(x)$ and a neighborhood $U(f)$ such that each function in U maps any point of V onto a point of W.

Definition. If t and t^* are two topologies for the same set of elements, and if every open set of t is an open set of t^*, we shall say that t is *stronger* than t^*, and shall write $t \subset t^*$.

Theorem 2. *If A is a locally compact Hausdorff space, and C is a class of continuous functions defined on A, then the k-topology is the strongest admissible topology that can be given to C.*

We refer the reader to [1, p. 482] for a proof of Theorem 2.

In particular, we can say that the group of homeomorphisms of a locally compact Hausdorff space can be given a strongest admissible topology.

To show the rôle of local compactness, we prove the following:

Given any admissible topology for the group of homeomorphisms H of the rational number system, one can construct another admissible topology for H which is not weaker than the first.

---

[11] This correspondence is also an isomorphism if C, and thus C^*, are groups of homeomorphisms.

[12] This proposition was suggested to us by G. Birkhoff.

Proof. Let H have an admissible topology. Then one can certainly find a neighborhood U of the identical homeomorphism and an interval V about 0 such that for $h \in U$ and $x \in V$ one will have $|h(x)| < 1$. Now let ξ be an irrational number lying between the endpoints of V. Construct a new topology for H by employing as subbasic neighborhoods sets of the form (F, W) where F is closed and contains no sequence (of rational numbers) converging to ξ, and W is open. This topology is clearly admissible. We shall show that none of the new neighborhoods of the identity, say $U^* = (F_1, \cdots, F_n; W_1, \cdots, W_n)$, is contained in U. For then surely we can find an open-closed interval surrounding ξ, and contained within V, which does not meet $F_1 \cup \cdots \cup F_n$; and a homeomorphism h can be set up which interchanges this open-closed interval with a corresponding neighborhood of $10 + \xi$, but leaves all other points fixed. Then $h \in U^*$ but h certainly does not lie in U.

4. Topologizing homeomorphism groups.

THEOREM 3. *The group of homeomorphisms of a locally compact Hausdorff space, becomes, with the g-topology, a topological group; and the g-topology is the strongest admissible topology for which this is true.*

Proof. We consider first the group H of homeomorphisms of a compact space A^*. In this case, the g- and k-topologies coincide, and the latter has just been proved to be the strongest admissible topology. Now let (K, W) be a neighborhood of a homeomorphism h. This means that $h(K) \subset W$, or that $h(K') \supset W'$, or that $h^{-1}(W') \subset K'$, whence (W', K') is a neighborhood of h^{-1}. Thus the process of set complementation in A^* shows that group-inversion transforms neighborhoods into neighborhoods, in H, and thus that the passage from h to h^{-1} is continuous.

Now let f and g be homeomorphisms and let a neighborhood of $h = fg$ (see [2]), (K, W), be given. This means that $f(g(K)) \subset W$, i.e., that $g(K) \subset f^{-1}(W)$. Since A^* is normal, we can find an open set V such that $g(K) \subset V$, $V^- \subset f^{-1}(W)$. Therefore (K, V) and (V^-, W) are neighborhoods of g and f, respectively, and it is easy to see that if g' belongs to the former, and f' to the latter, then $f'g'$ will send K into W, or $f'g' \in (K, W)$. Thus the group operations in H are all continuous: H is a topological group.

If H is the group of homeomorphisms of a locally compact Hausdorff space A, then let us apply Theorem 1 and consider the corresponding group H^* of homeomorphisms leaving the ideal point $I = A^* \cap A'$ fixed. Now H^* is clearly a closed subgroup of the group of all homeomorphisms of A^*, for if, for some h, $h(I) \neq I$, then (I, A) is a neighborhood of h not meeting H^*. There-

fore, by the preceding case, H^* is a topological group, and thus so also is H, with the g-topology.

Now suppose that we have another topology t^* for H, which is admissible and makes the passage h to h^{-1} bicontinuous in H. Let a g-neighborhood $U = (K, W)$ be given. If this is not already a k-neighborhood, $U^{-1} = (W', K')$ will be, and thus either U or U^{-1}, and hence both, will be open in the topology t^*, since the k-topology is the strongest admissible one. Therefore the topology t^* is weaker than the g-topology also.

REMARK. If in the definition of k-topology, we merely require that the set K be closed, we obtain a topology which makes the group of homeomorphisms of any normal space into a topological group. The proof resembles that of Theorem 3.

5. Homeomorphisms of locally connected spaces.

DEFINITION. A *locally connected topological space* is one in which the connected, open sets form a *complete system* of neighborhoods.

THEOREM 4. *If H is the group of homeomorphisms of a locally connected, locally compact Hausdorff space, and H is given the k-topology, then h^{-1} varies continuously with h in H, and H forms a topological group.*

Proof. The binary group operation will first be shown to be continuous; the local connectedness of the space A is not needed for this. Let (K, W) be a neighborhood of fg. This means that $g(K) \subset f^{-1}(W)$. Since A is locally compact, we can, at each point $x \in g(K)$, find an open set $V(x)$ whose closure is compact and contained in $f^{-1}(W)$. This covering by the V's can be reduced to a finite cover, and we have:

$$g(K) \subset V \subset V^- \subset f^{-1}(W), \text{ where } V = V(x_1) \cup \cdots \cup V(x_n),\ x_1, \cdots, x_n \in g(K).$$

Therefore (K, V) and (V^-, W) are neighborhoods of g and f, respectively, and if g' belongs to the former, and f' to the latter, then (as in Theorem 3) $f'g'$ must lie in (K, W).

Before proceeding to a discussion of h^{-1}, let us establish that the sets of the form (L, W), where L is compact, connected, and has a non-void interior, and W is open, constitute a subbase for the k-topology of H. Let $h \in (K, W)$ where K is any compact set. As W is locally connected, we can, for each $x \in K$, find a connected open set $V(x)$, whose closure is compact, such that $f(V(x)^-) \subset W$. Then $x_1, \cdots, x_n$ can be found such that $K \subset L_1 \cup \cdots \cup L_n$ where $L_i = V(x_i)^-$, which is connected. Clearly $(L_1, \cdots, L_n;\ W_1, \cdots, W_n) \subset (K, W)$, as desired.

Now let (L, W) be a k-neighborhood of f^{-1}, where L is compact, connected, and has a non-void interior $L'^{-\prime}$. As demonstrated in the earlier part of this proof, we can find an open set G such that $f^{-1}(L) \subset G \subset G^{-} \subset W$, where G^{-} is compact. We may similarly select an open set V containing G^{-}, with V^{-} compact and included in W. Then $f(G' \cap V^{-}) \subset L' \cap f(W)$. We can also find a point x for which $f(x) \in L'^{-\prime}$. Therefore $(x, G' \cap V^{-}; L'^{-\prime}, L' \cap f(W))$ is a k-neighborhood of f, which we shall refer to as U_0.

If $h \in U_0$, then $h(G' \cap V^{-}) \subset L' \cap f(W)$. By taking complements and reversing the inclusion sign, $L \cup f(W)' \subset h(G \cup V^{-\prime})$. Since $h(G)$ and $h(V^{-\prime})$ are disjoint open sets, and L is connected, L must lie in the one or the other. It cannot lie in $h(V^{-\prime})$ because $h(x) \in L'^{-\prime}$ and x is not in $V^{-\prime}$. Hence $L \subset h(G)$, or $h^{-1}(L) \subset G \subset W$. But this means that $h^{-1} \in (L, W)$. Thus inversion is continuous in H when it has the k-topology.

In other words, H forms a topological group.

6. Convergence in the k- and g-topologies.

DEFINITION. A directed system Δ is a partially ordered class with the property that if δ_1 and δ_2 are elements of Δ, then there is a $\delta \in \Delta$ for which $\delta > \delta_1$, and $\delta > \delta_2$.

DEFINITION. A function $\bar{x}$ defined on a directed system and taking its values in a topological space A is called a directed set. The value of $\bar{x}$ at $\delta \in \Delta$ is usally written x_δ and not $\bar{x}(\delta)$. When it causes no confusion, we shall call $\bar{x}$ "the directed set x_δ." x_δ will be said to converge to a limit $x \in A$ (in symbols, $x_\delta \rightarrow x$ or $\lim_{\delta \in \Delta} x_\delta = \lim_\delta x_\delta = x$) when for every neighborhood V of x, there is a $\delta_0 \in \Delta$ such that $\delta > \delta_0$ implies $x_\delta \in V$.

DEFINITION. Let Δ and N, and therefore $\Sigma = \Delta \times \mathrm{N}$, be directed systems.[13] Let x_δ be a directed set, and suppose that for each $\sigma = (\delta, \nu)$, $x'_\sigma = x_{\delta'}$ where $\delta' > \delta$. Then we shall call x'_σ a directed subset of x. The directed subset x'_σ will sometimes be denoted by the symbol $x_{\delta'}$.

The utility of this definition lies in the following lemma, which we shall occasionally use.

LEMMA 2. *If $x_\delta \rightarrow x$ and $x_{\delta'}$ is a directed subset of x_δ, then $x_{\delta'} \rightarrow x$. If x_δ is a directed set in a compact space A, then some directed subset of x_δ converges.*

Proof. Suppose that $x_\delta \rightarrow x$. Now, if a neighborhood $V(x)$ is given, then, for some δ_0, $\delta > \delta_0$ implies $x_\delta \in V$. Let $\sigma = (\delta, \nu)$ with any $\nu \in \mathrm{N}$. Then

[13] If $\sigma \in \Sigma$, then $\sigma = (\delta, \nu)$; $(\delta, \nu) > (\delta_0, \nu_0)$ means $\delta > \delta_0$ and $\nu > \nu_0$.

$\sigma > \sigma_0$ implies $x'_\sigma \in V$, whence $x_{\delta'} \to x$. Now suppose that $x_\delta \in A$, A being compact. Suppose that for each $x \in A$, there is a neighborhood $V(x)$ and a δ_x such that if $\delta > \delta_x$ then x_δ cannot lie in $V(x)$. Suppose that $A = V(x) \cup \cdots \cup V(z)$. Let $\delta > \delta_x, \cdots, \delta_z$. Then x_δ cannot lie in A. This contradiction shows that for some $x \in A$, and any δ and neighborhood V of x, then for some $\delta' > \delta$, $x_{\delta'} \in V$. The neighborhoods of x forming a directed system, we have $x_{\delta'} \to x$.

We can now consider convergence [14] properties of the k- and g-topologies.

THEOREM 5. *Let h_ν be a directed set in the group H of homeomorphisms of a locally compact Hausdorff space A. Then,*

i) *$h_\nu \to h$ in the k-topology if and only if $h_\nu(x_\delta) \to h(x)$ whenever $x_\delta \to x$ in A.*

ii) *$h_\nu \to h$ in the g-topology if and only if $h_\nu \to h$ and $h_\nu^{-1} \to h^{-1}$ in the k-topology.*

iii) *$h_\nu \to h$ in the g-topology if and only if $x_\delta \to x$ is always equivalent to $h_\nu(x_\delta) \to h(x)$ for any directed set x_δ in A.*

Proof. If $h_\nu \to h$ in the k-topology, then $h_\nu(x_\delta) \to h(x)$ whenever $x_\delta \to x$, because the k-topology is admissible. And if $h_\nu(x_\delta) \to h(x)$ whenever $x_\delta \to x$, one may introduce an admissible topology into the set $\{h, h_\nu\} = D$ by taking as the μ-th neighborhood of h, the set U_μ of all h_ν with $\nu > \mu$, and h. All $h_\nu \neq h$ shall be considered isolated. This patently gives D an admissible topology, and one in which $h_\nu \to h$. Since the k-topology is stronger than any admissible topology, $h_\nu \to h$ in the k-topology. This proves i).

For ii), suppose $h_\nu \to h$ and $h_\nu^{-1} \to h^{-1}$ in the k-topology. Let any subbasic neighborhood of h in the g-topology be given, say (K, W). If this is not already a k-neighborhood, then (W', K') must be a k-neighborhood of h^{-1}. Hence, for some ν_0, $\nu > \nu_0$ will imply $h_\nu^{-1}(W') \subset K'$, or $h_\nu(K) \subset W$, whence $h_\nu \to h$ in the g-topology. If (K, W) is a k-neighborhood, then for some ν_0, $\nu > \nu_0$ will imply $h_\nu(K) \subset W$, with the same result. If h_ν converges in the g-topology, it certainly converges in the k-topology. Thus we have ii).

iii) can be obtained by combining i) and ii).

---

[14] One of the most important properties, but one which we shall not require, is that for any directed set f_ν of continuous functions $f_\nu \to f$ in the k-topology if and only if $f_\nu(x)$ converges uniformly, on compact sets, to $f(x)$ [1]. The range space, of course, must be a metric, or at least a uniform, space.

An immediate application of this is a reformulation of the essential part of Theorem 4:

THEOREM 4′. *If A is a locally compact, locally connected Hausdorff space, and if h_ν is a directed set of homeomorphisms of A having the property that $h_\nu(x_\delta) \to x$ whenever $x_\delta \to x$, then $h_\nu^{-1}(x_\delta) \to x$ whenever $x_\delta \to x$.*

We shall give an example showing that local connectedness is not redundant here (nor in Theorem 4):

Let A be the set of real numbers $0, 1, 2, \cdots$ and $1/2, 1/3, \cdots$. Then A is locally compact, but not locally connected at 0. Consider the sequence $h_1, h_2, \cdots$ of homeomorphisms of A where

$$h_k(m) = \begin{cases} m \text{ if } m = 0, 1, 2, \cdots, k-1 \\ 1/k \text{ if } m = k \\ m-1 \text{ if } m = k+1, k+2, \cdots \end{cases}$$

$$h_k(1/m) = \begin{cases} 1/m \text{ if } m = 2, 3, \cdots, k-1 \\ 1/m+1 \text{ if } m = k, k+1, \cdots. \end{cases}$$

Since $1/m$ is the only converging sequence in A, we have only to note that $h_k(1/m) \to 0$ to see that h_k converges to the identity in the k-topology. But $h_k^{-1}(1/k) = k$, which diverges to ∞.

This shows also that the k-topology does not generally coincide with the g-topology, for h_k evidently does not converge in the latter.

7. Uniform structures in Spaces and Groups. Let A be any space, and let, as usual, $A \times A$ denote the class of all pairs (x, y), where $x, y \in A$. Let I denote the class of all pairs (x, x).

DEFINITION. A family $\mathfrak{A}$ of subsets of $A \times A$ is called a uniform structure for A if [Bourbaki 4, pp. 85, 86, and 92]

I. Each $R \in \mathfrak{A}$ contains I as a subset, and if $x \neq y$, then some $R \in \mathfrak{A}$ does not contain (x, y).

II. If R and S belong to $\mathfrak{A}$, then so does $R \cap S$.

III. If R and S belong to $\mathfrak{A}$, then there is an $S \in \mathfrak{A}$ such that $SS \subset R$.[15]

[15] By RS we mean the class of all pairs (x, z) for which there is an element $y \in A$ such that $(x, y) \in R$ and $(y, z) \in S$. We define S^n, inductively, $S^n = S(S^{n-1})$ and $S^2 = SS$. By $S(x)$ we mean the subset of A consisting of all z such that $(x, z) \in S \in \mathfrak{A}$. By R^{-1} we mean the class of all pairs in R, with the order reversed.

IV. $R^{-1} = R$, and the sets $R(x)$ form a complete system of neighborhoods for the point x in A.

For example, given any topological group H, one can define a uniform structure $\mathfrak{H}^*$ by saying that $(x, y) \in U^*$ if $xy^{-1} \in U$, and U is a neighborhood of the identity in H. Another uniform structure $^*\mathfrak{H}$ can be obtained by setting $(x, y) \in {}^*U$ if $x^{-1}y \in U$, U being any neighborhood of the identity in H. One may define a third uniform structure $^*\mathfrak{H}^*$ in which the relations are $^*U^* = {}^*U \cap U^*$, U^* and $^*U$ being defined as above [Weil, 15, p. 30].

A topological space A on which a uniform structure has been defined will be called a uniform space.

Definition. Let A have a uniform structure $\mathfrak{A}$, and let x_δ be a directed set in A such that for each $R \in \mathfrak{A}$ there is a δ_0 such that $\delta, \delta_1 > \delta_0$ implies $(x_\delta, x_{\delta_1}) \in R$. Then x_δ will be called a Cauchy directed set (relative to $\mathfrak{A}$). If every Cauchy directed set in A converges (has a limit in A), then A will be called complete (in the structure $\mathfrak{A}$).

A topological group is conventionally called complete if it is complete in $\mathfrak{H}^*$ or $^*\mathfrak{H}$, for if it is complete in one, it is likewise in the other (Weil, *loc. cit.*).

Theorem 6. *The g-topologized group H of homeomorphisms of a locally compact Hausdorff space is always complete in $^*\mathfrak{H}^*$, but not generally complete in $\mathfrak{H}^*$ and $^*\mathfrak{H}$.*

Proof. It will suffice to consider the k-topologized group of homeomorphisms of a compact Hausdorff space A, by Theorem 1, and the fact that the subgroup leaving a certain point fixed is always closed.

Suppose that h_ν is a Cauchy directed set relative to $^*\mathfrak{H}^*$. This means that $h_\nu^{-1}h_\mu \to$ identity and $h_\nu h_\mu^{-1} \to$ identity. For any $x \in A$, since A is compact, $h_\nu(x)$ will have a convergent directed subset $h_{\nu'}(x)$. Combining $h_{\nu'}(x) \to y$ with $h_\nu h_{\nu'}^{-1} \to$ identity by Theorem 5, i), we obtain the result that $h_\nu(x)$ converges for each x. Let its limit be called $h(x)$. We shall show that $h_\nu \to h$ in the k-topology. Suppose that $x_\delta \to x$. If $h_\nu(x_\delta) \not\to h(x)$, using the compactness of A, suppose that $h_{\nu'}(x_{\delta'}) \to y \neq h(x)$. By the previous reasoning, we have:

$$\lim_{\mu,\nu,\delta} h_\mu h_{\nu'}^{-1} h_{\nu'}(x_{\delta'}) = \lim_{\mu,\delta} h_\mu(x_{\delta'}) = y.$$

Now

$$\lim_\delta h_\mu(x_{\delta'}) = h_\mu(x) \text{ and } h_\mu(x) \to h(x).$$

Then

$$\lim_\mu \lim_\delta h_\mu(x_{\delta'}) = \lim_{\mu,\xi} h_\mu(x_{\delta'})$$

since the inner limit on the left, and the double limit on the right have both been shown to exist [Cf. Bourbaki, 4, p. 49]. This leads to $y = h(x)$, and this contradiction proves $h_\nu(x_\delta) \to h(x)$. Thus by Theorem 5, i), $h_\nu \to h$ in the k-topology. Let us also show that h is continuous. We now know that if $x_\delta \to x$, then $h_\nu(x_\delta) \to h(x)$. Therefore this double limit must equal the iterated limit

$$\lim_\delta \lim_\nu h_\nu(x_\delta) = \lim_\delta h(x_\delta)$$

since the inner limit exists, and hence $h(x_\delta) \to h(x)$. Because the condition on h_ν is equivalent to the same condition on h_ν^{-1}, we can conclude that $h_\nu^{-1}(x)$ also converges to a unique limit for each x, and defines a continuous function, $h^*(x)$. Applying Theorem 5, i) again, we have $h_\mu h_\nu^{-1}(x) \to h(h^*(x))$. But since, by hypothesis, $h_\mu^{-1}h_\nu \to$ identity, we can conclude $h(h^*(x)) = x$. Thus $h \in H$. This proves that H is complete in the structure $*\mathfrak{H}*$.

To show that H is not generally complete in $\mathfrak{H}^*$, consider the sequence h_n of homeomorphisms of the closed interval $[0, 2] = A$ where

$$h_n(x) = \begin{cases} tx, & 0 \leq x \leq 1 \\ (2-t)(x-1) + t, & 1 \leq x \leq 2, \end{cases} \text{, and } t = 1/n.$$

One can verify easily that $h_n h_m^{-1}$ converges uniformly to the identity, but $h_n(x) \to 0$ for each x in $[0, 1]$.[16]

Remarks on the compact metric case: If A has a metric, m, then the uniform structure $\mathfrak{H}^*$ corresponds to the well known metric m_1 in H, where $m_1(h, g) = \sup m(h(x), g(x))$, $x \in A$ [Cf. 3, p. 872], while $*\mathfrak{H}*$ corresponds to the metric m_2 where $m_2(h, g)$ is the larger of $m_1(h, g)$ and $m_1(h^{-1}, g^{-1})$, used by Oxtoby and Ulam [15, p. 879]. These metrics are of course equivalent (Cf. [15], *loc. cit.*) (perhaps this can be seen most quickly by noting that both determine the k-topology). This shows that although H is not complete in the sense of van Dantzig [5, p. 612], i. e., in m_1, it is of the second category, since it is complete in m_2.

8. Equicontinuous groups of homeomorphisms. The following is an obvious generalization of the concept of groups of isometries of a metric space.

Definition. An equicontinuous group H of homeomorphisms of a topological space A with a uniform structure $\mathfrak{A}$ is one such that for each $R \in \mathfrak{A}$ there is an $S \in \mathfrak{A}$ such that $(x, y) \in S$ implies $(h(x), h(y)) \in R$ for any $h \in H$ and any x, y in A.

---

[16] By using the remark of [14], one may prove in general that if $h_\nu \to f$, where f is continuous, then $h_\nu h_\mu^{-1} \to$ identity, in the k-topology.

LEMMA 3. *If H is an equicontinuous group of homeomorphisms of a uniform space A, then A can be given a uniformly equivalent* [15, p. 9] *structure $\mathfrak{A}^*$, which is invariant under H.*

Proof. For every $R \in \mathfrak{A}$, let R^* be the class of all pairs (u, v) such that for some $h \in H$ and some $(x, y) \in R$, $u = h(x)$ and $v = h(y)$. Since H has an identity element, $R \subset R^*$. But as H is equicontinuous, there is an S such that $(x, y) \in S$ implies $(h(x), h(y)) \in R$ for every h in H, whence $S^* \subset R \subset R^*$. This makes $\mathfrak{A}^*$ uniformly equivalent to $\mathfrak{A}$. If $(u, v) \in R^*$, then, for some x, y, h, we have $(x, y) \in R$ and $u = h(x)$ and $v = h(y)$, and so for any $g \in H$, $(g(u), g(v)) = (gh(x), gh(y)) \in R^*$. Therefore H sends R^* into R^*, and as H is a group, $\mathfrak{A}^*$ must be invariant.

COROLLARY. *If H is an equicontinuous group of homeomorphisms of a metric space A, then A can be given a uniformly equivalent metric which makes H a group of isometries.*[17]

This follows from the fact that a countable uniform structure determines a uniformly equivalent metric, and conversely [Weil, *op. cit.*, p. 16].

The following is a generalization of a theorem of van der Waerden and van Dantzig [14, p. 374].

THEOREM 7. *Any equicontinuous group H of homeomorphisms with the g-topology, of a locally compact, connected uniform space A is dense in some (equicontinuous) locally compact topological group H^- of homeomorphisms of A. In H^-, h_ν converges in the g-topology if and only if $h_\nu(x)$ converges for each x in A.*

Proof. Consider the set H^- of all functions on A into itself which are pointwise limits of directed sets in H. We shall prove that H^- has the properties set forth. Let us show first that H^- contains only homeomorphisms.

Suppose that $f \in H^-$; then some directed set h_ν converges pointwise to f. Suppose that $R \in \mathfrak{A}$, and that $S \in \mathfrak{A}$ is such that $(x, y) \in S$ implies $(h(x), h(y)) \in R$, and suppose that $(x, y) \in S$. Hence $(h_\nu(x), h_\nu(y)) \in R$ and so in the limit $(f(x), f(y)) \in R^3$ (see [15]). This shows that the whole set H^- is equicontinuous, and in particular, that f is continuous.

Given a relation R as above, and using the same $S \in \mathfrak{A}$, we can find a ν_0 such that $\nu > \nu_0$ implies $(h_\nu(x), f(x)) \in S$, and hence that $(x, h_\nu^{-1}f(x)) \in R$. Thus f has an inverse f^{-1} to which h_ν^{-1} converges pointwise, and which is therefore continuous. We need therefore only to prove that $f(A) = A$, and for

[17] Cf. Eilenberg, *op. cit.*, p. 79, § 8.

this we interpolate a lemma. In it we suppose that $\mathfrak{A}$ is invariant under H, which can always be done by Lemma 3.

LEMMA 4. *If $h_\nu \in H$ is a directed set, and $a \in A$, and $R^4(a)^-$ is compact, and if $h_\nu(b)$ converges for some $b \in R(a)$, then for each $x \in R(a)$, $h_\nu(x)$ has a convergent directed subset $h_{\nu'}(x)$.*

Proof. Select μ such that if $\nu > \mu$ then $(h_\nu(b), h_\mu(b)) \in R$. If $(x, a) \in R$, then $(h_\nu(x), h_\nu(a)) \in R$. This is true in particular if $x = b$ and also then for $\nu = \mu$. This gives four conditions which can be combined to $(h_\nu(x), h_\mu(a)) \in R^4$, or $h_\nu(x) \in R^4(h_\mu(a)) = h_\mu(R^4(a))$; and this last set has a compact closure because it is homeomorphic to $R^4(a)$. Hence $h_\nu(x)$ has a convergent directed subset, and Lemma 4 is proved.

Now $h_\nu^{-1}(b)$ converges at some point $b \in f(A)$. Let R, x, and a have the meaning they have in Lemma 4. Such an R exists for a given a because A is locally compact. Then $h_\nu^{-1}(x)$ has a convergent directed subset which converges to a point, say $h_{\nu'}^{-1}(x) \rightarrow c$.

Let $S \in \mathfrak{A}$ be given, and select $T \in \mathfrak{A}$ such that $T^3 \subset S$. We can find a μ such that $\nu > \mu$ will imply $(h_\nu(c), f(c)) \in T$, and imply $(h_{\nu'}(c), h_\nu(c)) \in T$, and imply $(h_{\nu'}^{-1}(x), c) \in T$ or $(x, h_{\nu'}(c)) \in T$. This gives $(f(c), x) \in S$, and therefore $f(c) = x$.

Therefore x as well as b lies in $f(A)$, which is therefore open. On the other hand, the point a might have been taken as any point in $f(A)^-$, and x taken as coinciding with a. Thus $f(A)$ is also closed. Since it is not void, and since A is connected, $f(A) = A$.

Thus G^- contains only homeomorphisms.

Next we prove that convergence in the g-topology follows from pointwise convergence in H^-. We already know that if $h_\nu \rightarrow h$ pointwise, then $h_\nu^{-1} \rightarrow h^{-1}$ pointwise. Therefore we need only check that $h_\nu \rightarrow h$ in the k-topology, by Theorem 5.

Suppose that $x_\delta \rightarrow x$, and let $R \in \mathfrak{A}$ be given. Select $S \in \mathfrak{A}$ such that $S^5 \subset R$, and ν_0 and δ_0 such that $\nu > \nu_0$, $\delta > \delta_0$ implies $(x_\delta, x) \in S$ and $(h_\nu(x), h(x)) \in S$. As we have seen before, we can be sure that $(h_\nu(x_\delta), h_\nu(x)) \in S^3$ when $(x_\delta, x) \in S$, and so $(h_\nu(x_\delta), h(x)) \in R$, proving $h_\nu \rightarrow h$ in the k-topology.

Therefore H^- is a closed subgroup of the g-topologized group of all homeomorphisms of A, and thus is a topological group.

It remains to show that H^- is locally compact. Let $f \in H^-$ and select any point $a \in A$. Suppose that $R(f(x))^-$ is compact. We shall show that the neighborhood $U = (a, R(f(a)))$ of f has a compact closure. If h_ν is a

directed set in U, then $h_\nu(a)$ is a directed set in $R(f(a))$, and hence has a convergent directed subset, $h_{\nu'}(x)$. The argument following the proof of Lemma 4 may be applied to show that $h_{\nu'}(x)$ converges for every x. Thus U has a compact closure.

REMARK. H^- is easily seen to be compact if A is.

COROLLARY. (van der Waerden and van Dantzig, [14, p. 374]).[18] *The group of isometries of a locally compact metric space which contains but a finite number of connected components forms a locally compact topological group.*

9. Quotient Spaces. In this section, we shall develop some topological connections between groups of homeomorphisms and the regions over which they are transitive.

Suppose that H is a group of homeomorphisms of a topological space A, that b is a fixed point of A, that ${}_bH$ is the subgroup of H consisting of homeomorphisms which leave b invariant, and finally that $H(b)$ is the set of all points into which H can transform b,—otherwise known as the orbit of b (under H). Then there is a natural $1-1$ correspondence between cosets of ${}_bH$ and the points of $H(b)$ (Cf. van Dantzig [5, p. 610]),

$$\phi : \phi({}_bHg) = g(b).$$

This correspondence is possible because $g(b)$ depends only on the coset of ${}_bH$ in H in which the homeomorphism g lies.

LEMMA 5. *If a group H of homeomorphisms of a topological space A has an admissible topology, and the natural topology* (Cf. [12, p. 58]) *be introduced into the quotient space $Q = H/{}_bH$, then ϕ is a continuous mapping.*

Proof. ${}_bH$ is certainly closed, since the topology of H is admissible. Then the natural topology in $Q = H/{}_bH$ is as follows: Let U be an arbitrary open set in H and let ${}_bHU$ denote the set of all cosets ${}_bHg$ where $g \in U$. Then these sets ${}_bHU$ are taken as a basis in Q. Now let an element ${}_bHg$ of Q be given, and a neighborhood W of $\phi({}_bHg) = g(b)$ in A. Now (b, W) is a k-neighborhood of $g \in H$, and thus an open set in H (Cf. Theorem 2). Clearly, if $h \in (b, W)$, then $\phi({}_bHh) = h(b) \in W$. Thus ϕ is continuous.

LEMMA 6. *In order that ϕ^{-1} be continuous, it is necessary and sufficient that for each neighborhood U of the identity in H, and for each $b \in A$, the set $U(b) \subset A$ (that is, the orbit of b under U) have a non-void interior, relative to $H(b)$, which contains b.*

---

[18] The condition of separability assumed by van Dantzig and van der Waerden follows from the remaining conditions (Cf. Sierpinski [13, p. 111]).

Proof. Let $x \in H(b) \subset A$ be given, and also a neighborhood ${}_bHUg$ of the coset ${}_bHg \in Q$ be given. Then $\phi({}_bHUg) = U(x)$ and by hypothesis there is an open set W in A such that $x \in W \cap H(b) \subset U(x)$. Obviously $\phi^{-1}(W) \subset {}_bHUg$.

Conversely, suppose that ϕ^{-1} is continuous. Then $\phi({}_bHUg) = U(g(b))$ is an open set (relative to $H(b)$).

Combining these, we infer

THEOREM 8. *If H is an admissibly topologized group of homeomorphisms of a topological space A, and if for every neighborhood U of the identity in H, we have $U(b)$ open relative to $H(b)$, for b in A, then the quotient space $H/{}_bH$ is homeomorphic to the orbit $H(b)$.*

COROLLARY. *If H is a locally compact, admissibly topologized group of homeomorphisms of a topological space A, and if H has a denumerable basis while the orbit $H(b)$ is not the sum of countably many sets which are nowhere dense on $H(b)$, then the quotient space $H/{}_bH$ is homeomorphic to the orbit $H(b)$.*[19]

Proof. Let U be any neighborhood of the identity in H, and let V be a neighborhood of the identity whose closure is compact and contained in U. Now H can be covered by a countable set $V_1, V_2, V_3, \cdots$ of neighborhoods each a translation of V, since it has a countable basis, and so the orbit $H(b)$ is covered by the sets $\bar{V}_1(b), \bar{V}_2(b), \cdots$, which are all homeomorphic to $\bar{V}(b)$, which is closed, and hence nowhere dense if it has no interior (all relative to the orbit $H(b)$). Now select V so that $V^{-1}V \subset U$ (Cf. [12, p. 54]). Then there is an open set W, and a point x in $H(b)$ such that $x \in W \cap H(b) \subset V(b)$. Suppose $x = g(b)$, where $g \in V$. Then $b \in g^{-1}(W \cap H(b)) \subset g^{-1}V(b) \subset U(b)$. Thus the hypothesis of Lemma 6 is fulfilled.

We shall now give an example of a compact metric space A with a locally compact group of homeomorphisms H for which the hypothesis and conclusion of Theorem 8 fail to be fulfilled.

We use cylindrical coordinates (r, θ, z) to describe A in Cartesian 3-space. Let S_n be the closed segment connecting $p_n = (1, n, 0)$ with $q_n = (0, 0, 1 - 2^{-n})$ when $n \geq 1$, and connecting $p_n = (1, n, 0)$ with $q_n = (0, 0, 2^n - 1)$ when $n \leq -1$, and let S_0 connect $p_0 = (1, 0, 0)$ with the origin $q_0 = (0, 0, 0)$. It is essential that θ be measured in radians so that the points $(1, n, 0)$ shall be all distinct and dense on the circle $r = 1$ in the (x, y)-plane.

---

[19] This covers the important case in which H is a transitive Lie group of transformations. The proof is an obvious modification of [12, p. 65] (Theorem 13).

Then A is to be the union of all the sets S_n together with the segment connecting $(0, 0, -1)$ with $(0, 0, 1)$ on the z-axis, and the double conical mantle enclosing all the segments whose equation can be written as $|z| + r = 1$, r being always ≥ 0.

It is readily seen that A is closed, and hence compact. But now observe that a homeomorphism h which leaves the end p_n of a segment S_n fixed must leave the other end q_n fixed also. For each n, a homeomorphism h_n can be constructed which rotates the mantle of A through n radians, and sends each S_m into S_{m+n}, p_m into p_{m+n}, q_m into q_{m+n} and acts linearly on the segment connecting q_m with q_{m+1}. Then $H = \{h_n\}$ is a group homeomorphic to the group of integers. Only the identity leaves p_0 fixed, and yet the orbit of p_0, viz. $(\cdots, p_{-1}, p_0, p_1, p_2, \cdots)$ is dense on itself (hence, by Theorem 8, Corollary, must be the sum of a countable number of nowhere dense subsets of itself). Nor can this situation be patched up by choosing a different topology for H, for the orbit of q_0 is homeomorphic to the set of integers.[20]

10. The topology of Lie groups. The question arises: suppose we take a Lie group of transformations, G, and, ignoring the topology of G inherited from its parameter space, introduce the g-topology into G, then how does this topology compare with the original? We shall first introduce the concept of a Lie group of transformations, limiting ourselves to the analytic case (Pontrjagin, [12, p. 287]): An r-dimensional Lie group G is a topological group for which some neighborhood of the identity can be homeomorphically mapped on a neighborhood of Cartesian r-space in such a way that the group operations in G are analytic when expressed in terms of the coordinates $(a_1, \cdots, a_r)$, called parameters, thus introduced into this neighborhood of the identity. G is a Lie group of transformations if each $g \in G$ determines a bicontinuous transformation T of an analytic manifold A, such that if coordinates $(x_1, \cdots, x_n)$ are chosen locally in A, then this transformation T can be expressed by $\bar{x}_i = f_i(x_1, \cdots, x_n; a_1, \cdots, a_n)$ where f_i is continuous in all its arguments simultaneously, and analytic in each.[21] One also demands that T leave each point of A fixed if and only if it corresponds to the identity in G.

[20] Even under the group of all homeomorphisms of A, the orbits of p_0 and q_0 are not homeomorphic, being the same as under H. And yet the subgroup leaving p_0 invariant coincides with that leaving q_0 fixed. Thus we have a counterexample to a theorem of van Dantzig [5, p. 610], which says, (using the terminology of our Lemma 5) that if A is locally compact and H contains all homeomorphisms of A, then ϕ is a homeomorphism.

[21] One may define all this with functions s times continuously differentiable, rather than analytic. Theorem 9 would require only $s \geqq 2$.

THEOREM 9. *Suppose that G is a Lie group of transformations. Let the topology in G be replaced by the g-topology, giving a topological group H of transformations. Then, if G is a compact Lie group, H and G are topologically isomorphic, and, in general, when G is locally compact, the maximal connected subsets of the open sets of H form a basis for G.*

Proof. In the first place, the topology of G is always admissible, and therefore weaker than the k- or g-topologies, which are equivalent here, of course. Therefore, if G is compact, H and G must be homeomorphic (see [12, p. 65]).

For the general case where G is locally compact, we suppose that a neighborhood U of the identity in G be given, such that U^- is compact.

Since, by definition, no element on $U^- \cap U'$ leaves all points of A fixed, and since $U^- \cap U'$ is compact, we must be able to find a finite set of points $p_1, \cdots, p_m$ in A together with open sets $V_1, \cdots, V_m$ where $p_k \in V_k$, $k = 1, \cdots, m$, such that if g lies on $U^- \cap U'$ then the corresponding T drives some p_k out of its V_k, for the following reason: the subgroup ${}_pG$ of G leaving $p \in A$ fixed is closed; therefore some finite set of such subgroups must have a void intersection on the set $U^- \cap U'$. This gives the points $p_1, \cdots, p_m$. To get the $V_1, \cdots, V_m$, one observes that the sum of the distances the p_k are moved by g on the compact set $U^- \cap U'$ is bounded away from zero by a positive number d, and that spheres of radius d/m about the p_k will serve.

It is not hard to see that if U is sufficiently small then this means that if the matrix M is composed by placing m blocks $B_1, B_2, \cdots, B_m$ together

$$M = \| B_1 \, B_2 \cdots B_m \|$$

where

$$B_k = \begin{Bmatrix} \dfrac{\partial f_1(p_k; a)}{\partial a_1} & \cdots & \dfrac{\partial f_n(p_k; a)}{\partial a_1} \\ \vdots & \cdots & \vdots \\ \dfrac{\partial f_1(p_k; a)}{\partial a_r} & \cdots & \dfrac{\partial f_n(p_k; a)}{\partial a_r} \end{Bmatrix}$$

then M will have a minor or rank r which does not vanish on U.

This means that we can solve for the $(a_1, \cdots, a_r)$ in terms of some set of r coordinates of the transforms $\bar{p}_k$ of the p_k, this inversion:

$$a_i = \phi_i(\bar{p}_1, \cdots, \bar{p}_m)$$

being continuous and valid for $\bar{p}_k \in V_k$, where perhaps the V_k must be diminished in radius.

Now suppose that L is a connected set in H lying in the k-neighborhood $(p_1, \cdots, p_m;\ V_1, \cdots, V_m)$, and containing the identity in H. Then $L(p_k) \subset V_k$, and hence, by the inversion $a_i = \phi_i(\bar{p}_1, \cdots, \bar{p}_m)$, determines a connected set S in G, containing the identity. If S met the set $U^- \cap U'$, some p_k would be driven out of its V_k, whence $S \subset U$. This means that the connected component of $(p_1, \cdots, p_m;\ V_1, \cdots, V_m)$ lies in U, thus proving the theorem.

REMARK. The part of this theorem relating to compact Lie groups of transformations can be extended to any compact, effective (i.e., only the identity of the group leaves all points of A fixed) group of transformations, provided its topology is admissible (this is sometimes included in the concept "transformation group" (Cf. [9, p. 365])), since the proof for that part involves only the admissibility and compactness of the topology.

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

BIBLIOGRAPHY.

[1] R. F. Arens, "A topology for spaces of transformations," *Annals of Mathematics*, vol. 47 (1946), pp. 480-495.

[2] S. Banach, *Théorie des opérations linéaires*, Warsaw (1932).

[3] G. Birkhoff, "The topology of transformation sets," *Annals of Mathematics*, vol. 35 (1934), pp. 861-875.

[4] N. Bourbaki, *Topologie Générale*, Paris (1940).

[5] D. van Dantzig, "Zur topologischen Algebra, I. Komplettierungstheorie," *Mathematische Annalen*, vol. 107 (1933), pp. 587-626.

[6] S. Eilenberg, "Sur les groupes compacts d'homéomorphies," *Fundamenta Mathematicae*, vol. 28 (1937), pp. 75-80.

[7] R. H. Fox, "On topologies for function spaces," *Bulletin of the American Mathematical Society*, vol. 51 (1945), pp. 429-432.

[8] S. Lefschetz, *Algebraic topology*, American Mathematical Society Colloquium Publications, vol. 27 (1942).

[9] D. Montgomery, "Compact groups of transformations each of which is analytic," *Annals of Mathematics*, vol. 46 (1945), pp. 365-371.

[10] D. Montgomery and H. Samelson, "Transformation groups of spheres," *Annals of Mathematics*, vol. 44 (1943), pp. 454-470.

[11] J. C. Oxtoby and S. M. Ulam, "Measure preserving homeomorphisms and metric transitivity," *Annals of Mathematics*, vol. 42 (1941), pp. 874-920.

[12] L. Pontrjagin, *Topological Groups*, Princeton University Press (1939).

[13] W. Sierpinski, "Sur les espaces métriques localement séparables," *Rundamenta Mathematicae*, vol. 21 (1933), pp. 107-113.

[14] B. L. van der Waerden and D. van Dantzig, "Über metrisch homogene Räume," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, vol. 6 (1928), pp. 367-376.

[15] A. Weil, "Sur les Espaces à Structure Uniforme" . . . , *Actualités Scientifiques et Industrielles*, no. 551, Paris, 1937.

ON A CLASS OF PARTIAL DIFFERENTIAL EQUATIONS OF EVEN ORDER.*

By JOAQUIN B. DIAZ.

1. Introduction.[1] In this paper, we consider linear partial differential equations of the form

$$L^N u = 0, \tag{1.1}$$

where $u = u(x, y)$ and L is a certain linear partial differential operator of the second order. The most important special case is that of

$$Lu = \Delta u + \phi(x) u_x + \psi(y) u_y, \tag{1.2}$$

where $\Delta \equiv \partial^2/\partial x^2 + \partial^2/\partial y^2$; then (1.1) becomes

$$L^N u = \Delta^N u + \text{lower order terms} = 0.$$

Our main purpose is to construct a sequence of particular solutions of (1.1) having the following properties: (a) In the neighborhood of each point at which a solution of the equation $L^N u = 0$ is analytic, this solution can be expanded in a uniformly and absolutely convergent series of linear combinations of these particular solutions; (b) Each particular solution is defined and analytic in the whole plane and can be obtained by quadratures from the coefficients of L. Furthermore, all of these particular solutions are linear combinations of certain functions independent of N.

Our method is based upon the solution of partial differential equations with constant coefficients by the method of hypercomplex variables. [See Scheffers (3) [2], and Ketchum (2); Ward (5) gives an exhaustive bibliography on analytic functions of linear algebras.] This method involves replacing the given partial differential equation of higher order by a system of partial differential equations of the first order. Sobrero (4) has treated in detail the equation $\Delta^2 u = 0$ from this standpoint. For the equation

$$\Delta^N u = 0 \tag{1.3}$$

the procedure is as follows. Formally, we determine the hypercomplex

* Received September 13, 1945.

[1] Presented to the American Mathematical Society, September 17, 1945. The problem was suggested to the author by Dr. Lipman Bers.

[2] Numbers in parenthesis refer to papers in the bibliography.

unit j_N so that every function $u = f(x,y) = f(x + j_N y)$ shall be a solution of (1.3). By direct substitution, setting $f^{(2N)} = \dfrac{d^{2N}f}{d^{2N}(x + j_N y)}$, we obtain

$$\Delta^N f = \sum_{l=0}^{N} \binom{N}{l} \frac{\partial^{2N} f}{\partial x^{2(N-l)} \partial y^{2l}} = f^{(2N)} \sum_{l=0}^{N} \binom{N}{l} j_N^{2l}.$$

Thus, the necessary and sufficient condition that every function $f(x,y) = f(x + j_N y)$ be a solution of (1.3) is that

$$(1.4) \qquad (1 + j_N^2)^N = 0.$$

From (1.4), it is seen that any power of j_N greater than the $2N$-th may be expressed as a linear combination of $1, j_N, \cdots, j_N^{2N-1}$. The hypercomplex valued function $f(x,y) = \sum_{i=0}^{2N-1} a_i(x,y) j_N^i$ is said to be analytic at (x_0, y_0) provided there exists a neighborhood of (x_0, y_0) such that if (x_1, y_1) is any point of this neighborhood, the difference quotient

$$\frac{f(x,y) - f(x_1,y_1)}{(x - x_1) + j_N(y - y_1)}$$

approaches a limit independent of the mode of approach of the point (x,y) to the point (x_1, y_1). Letting (x,y) approach (x_1,y_1) in the directions parallel to the x and y axes, and equating the results, we obtain

$$\partial f/\partial x = (1/j_N)(\partial f/\partial y);$$

that is

$$\sum_{i=0}^{2N-1} a_{i,x} j_N^i = (1/j_N) \sum_{i=0}^{2N-1} a_{i,y} j_N^i,$$

where $a_{i,x} = \partial a_i/\partial x$; etc. Multiplying through by j_N in the last equation and employing (1.4) in the form

$$j_N^{2N} = -\sum_{i=0}^{N-1} \binom{N}{i} j_N^{2i},$$

we obtain

$$\sum_{k=0}^{N-1} \left[a_{2k-1,x} - a_{2N-1,x} \binom{N}{k}\right] j_N^{2k} + \sum_{k=0}^{N-1} a_{2k,x} j_N^{2k+1}$$
$$= \sum_{k=0}^{N-1} a_{2k,y} j_N^{2k} + \sum_{k=0}^{N-1} a_{2k+1,y} j_N^{2k+1}.$$

To simplify the writing, we set $a_{-1} \equiv 0$. This device of introducing zero terms will be used throughout, without explicit mention. Equating coefficients of like powers of j_N, we see that

$$a_{2k,x} = a_{2k+1,y},$$
$$(1.5) \qquad a_{2k,y} = a_{2k-1,x} - a_{2N-1,x}\binom{N}{k},$$

$(k = 0, \cdots, N-1)$. The system (1.5) bears the same relation to (1.3) as the Cauchy-Riemann equations bear to Laplace's equation. To complete the analogy, it remains to show that each of the $2N$ functions a_i satisfies (1.3), and that given any solution u of (1.3), there is a solution a_i, $i = 0, \cdots, 2N-1$, of (1.5) such that $a_0 = u$.

Returning now to (1.2), we rewrite L in the form

$$(1.6) \qquad Lu = (\tau/\sigma)[(\sigma u_x/\tau)_x + (\sigma u_y/\tau)_y],$$

where

$$\sigma = \exp\left(\int \phi dx\right), \qquad \tau = \exp\left(-\int \psi dy\right).$$

Actually, we shall consider the more general case

$$(1.7) \qquad Lu = (\tau_1/\sigma_2)[(\sigma_1 u_x/\tau_1)_x + (\sigma_2 u_y/\tau_2)_y].$$

If σ_i and τ_i do not change sign, the σ_i's have the same sign, and the τ_i's have the same sign, then (1.7) can be transformed into (1.6) by the change of variables

$$\xi = \int (\sigma_2/\sigma_1)^{1/2} dx, \qquad \eta = \int (\tau_2/\tau_1)^{1/2} dy.$$

We prefer to treat the general case because no additional difficulties arise and the formal part of our considerations remains valid even when the σ's and τ's change sign. In particular, this is true when L is of hyperbolic or of mixed type.

The system of first order equations connected with (1.1) is

$$\sigma_1(x) a_{2k,x} = \tau_1(y) a_{2k+1,y}$$
$$(1.8) \qquad \sigma_2(x) a_{2k,y} = \tau_2(y)\left[a_{2k-1,x} - a_{2N-1,x}\binom{N}{k}\right],$$

$(k = 0, \cdots, N-1)$, which arises in a natural way from (1.5).

Bers and Gelbart (1) have studied the system of equations [3]

---

[3] A system of equations of the form (1.9), with $\sigma_1 = \sigma_2 = \tau_1 = 1$, and which appears in the theory of compressible fluids, has been discussed independently by Professor S. Bergman, who introduced an operator analogous to the "formal powers" employed by Bers and Gelbart [1]. See S. Bergman, "The hodograph Method in the Theory of Compressible Fluids" (Supplement to R. V. Mises and K. O. Friedrichs,

$$\begin{aligned} &\sigma_1(x)u_x = \tau_1(y)v_y, \\ &\sigma_2(x)u_y = -\tau_2(y)v_x, \end{aligned} \tag{1.9}$$

which is obtained from the Cauchy-Riemann equations just as (1. 8) is obtained from (1. 5). By eliminating v from (1. 9), we obtain the equation $Lu = 0$.

To simplify the writing, we use the following real functions

$$\begin{aligned} X(x) &= \int_0^x (1/\sigma_1(\xi))\,d\xi, \qquad Y(y) = \int_0^y (1/\tau_1(\eta))\,d\eta, \\ X^*(x) &= \int_0^x \sigma_2(\xi)\,d\xi, \qquad Y^*(y) = \int_0^y \tau_2(\eta)\,d\eta, \end{aligned} \tag{1.10}$$

and rewrite (1. 8) in the form

$$\begin{aligned} &a_{2k,X} = a_{2k+1,Y}, \\ &a_{2k,Y^*} = a_{2k-1,X^*} - a_{2N-1,X^*}\binom{N}{k}, \end{aligned} \tag{1.11}$$

$(k = 0, \cdots, N-1)$, where $a_{2k,X} = \partial a_{2k}/\partial X$, etc. We shall refer to (1. 8), or (1. 11), as the system $E(\Sigma, N)$, where Σ is the matrix of the coefficients of (1. 8)

$$\Sigma = \left\| \begin{matrix} \sigma_1 & \tau_1 \\ \sigma_2 & \tau_2 \end{matrix} \right\|. \tag{1.12}$$

We remark that if $u = u(x, y)$ has continuous second derivatives, then $\partial^2 u/\partial X \partial Y = \partial^2 u/\partial Y \partial X$, but that in general $\partial^2 u/\partial X \partial X^* \neq \partial^2 u/\partial X^* \partial X$, etc. With the aid of (1. 10), the operator L of (1. 7) may be rewritten in the form

$$Lu = u_{XX^*} + u_{Y^*Y}.$$

Together with (1. 8), we shall be led to consider the system

$$\begin{aligned} &(1/\sigma_2(x))a_{2k,x} = \tau_1(y)a_{2k+1,y}, \\ &(1/\sigma_1(x))a_{2k,y} = \tau_2(y)\left[a_{2k-1,x} - a_{2N-1,x}\binom{N}{k}\right], \end{aligned} \tag{1.13}$$

$(k = 0, \cdots, N-1)$, which may be rewritten in the form

$$\begin{aligned} &a_{2k,X^*} = a_{2k+1,Y}, \\ &a_{2k,Y^*} = a_{2k-1,X} - a_{2N-1,X}\binom{N}{k}. \end{aligned} \tag{1.14}$$

Fluid Dynamics), mimeographed lecture notes, Brown University, 1942, especially pages 23-24; and "A Formula for the Stream Function of Certain Flows," *Proceedings of the National Academy of Sciences*, vol. 29 (1943), pp. 276-281, by the same author.

$(k=0,\cdots,N-1)$. We shall refer to (1.13), or (1.14), as the system $E(\Sigma',N)$, where Σ' is the matrix of the coefficients of (1.13)

$$\Sigma' = \left\| \begin{matrix} (1/\sigma_2) & \tau_1 \\ (1/\sigma_1) & \tau_2 \end{matrix} \right\|. \tag{1.15}$$

It will be assumed throughout that the functions $\sigma_1(x)$, $\sigma_2(x)$, $\tau_1(y)$, and $\tau_2(y)$ are *positive analytic functions defined for all values of their respective real variables*, but it is clear that some of the results hold under weaker hypotheses. All the hypercomplex valued functions we shall deal with are defined on the (x,y) plane. Accordingly, we shall write $f(z)=\sum_{i=0}^{2N-1} a_i j_N{}^i$ for a typical function, where $z=(x,y)=x+iy$ is a variable point of the (x,y)-plane and $a_i=a_i(x,y)$, and speak of the domain in the plane on which f is defined. By $|z|$ is understood the distance of the point z from the origin and by a domain is understood an open connected set.

In order to construct a function theory for (1.11), it is first necessary to study the algebra generated by the unit j_N of (1.4). This is done in Section **2**. The real algebra of order $2N$ generated by j_N is shown to be isomorphic to a certain complex algebra of order N; and a criterion for a number of the algebra to possess an inverse is given. In Section **3** (Σ,N)-monogenic functions are introduced. These are functions $f(z)=\sum_{i=0}^{2N-1} a_i j_N{}^i$ such that the a_i satisfy the system $E(\Sigma,N)$, and they take the place of analytic functions. For these functions, the processes of (Σ,N)-differentiation and (Σ,N)-integration are defined. These two processes transform (Σ,N)-monogenic functions into (Σ',N)-monogenic functions [see (1.14) and (1.15)]. The process of "contraction," which transforms (Σ,N)-monogenic functions into $(\Sigma,N-M)$-monogenic functions, is treated in Section **4**. The relations between (Σ,N)-, (Σ',N)-, $(\Sigma_,,N)$-, and $(\Sigma_,',N)$-monogenic functions are discussed in Section **6**, where the matrices of coefficients $\Sigma_,$ and $\Sigma'_,$ are defined. In Section **5** it is proved that if $f(z)$ is (Σ,N)-monogenic, then the real functions a_{2k} satisfy the equation $L^N u=0$. The process of (Σ,N)-integration enables us to construct special (Σ,N)-monogenic functions, called "formal powers," which bear the same relation to $E(\Sigma,N)$ as the powers of $x+iy$ bear to the Cauchy-Riemann equations. By means of these "formal powers," it is possible (Section **8**) to characterize completely all (Σ,N)-monogenic functions for which one of the functions a_i vanishes identically ("null" functions). It is then proved (Section **9**) that every analytic function satisfying $L^N u=0$ is an "even component" of a (Σ,N)-

monogenic function determined to within a "null function." The main result concerning (Σ, N)-monogenic functions is the expansion theorem proved in Section **10**, which states that a (Σ, N)-monogenic function may be expanded uniquely in a formal power series in the neighborhood of any point at which the function is (Σ, N)-monogenic. Using this expansion theorem, we construct the sequence of particular solutions of $L^N u = 0$ mentioned at the outset.

2. The algebras A_{2N}.

A. For $N = 1, 2, 3, \cdots$ let A_{2N} denote the algebra of order $2N$ over the real field, whose elements are the polynomials

$$\sum_{i=0}^{2N-1} a_i j_N{}^i, \tag{2.1}$$

where the a_i are real numbers and the unit j_N satisfies the equation

$$(1 + j_N{}^2)^N = 0. \tag{2.2}$$

Equality is defined by postulating that

$$\sum_{i=0}^{2N-1} a_i j_N{}^i = \sum_{i=0}^{2N-1} b_i j_N{}^i,$$

if, and only if, $a_i = b_i$ for $n = 0, \cdots, 2N - 1$.

Addition is defined by the identity

$$\sum_{i=0}^{2N-1} a_i j_N{}^i + \sum_{i=0}^{2N-1} b_i j_N{}^i = \sum_{i=0}^{2N-1} (a_i + b_i) j_N{}^i.$$

Multiplication is defined by the identity

$$\left(\sum_{i=0}^{2N-1} a_i j_N{}^i\right)\left(\sum_{i=0}^{2N-1} b_i j_N{}^i\right) = \sum_{i=0}^{4N-2} \left(\sum_{l+k=i} a_l b_k\right) j_N{}^i = \sum_{i=0}^{2N-1} c_i j_N{}^i,$$

where the c_i's are obtained by employing the relation [see (2.2)]

$$j_N{}^{2N} = -\sum_{k=0}^{N-1} \binom{N}{k} j_N{}^{2k}. \tag{2.3}$$

For each N, the system possesses the identity element 1. If a number of A_{2N} has an inverse, then the inverse is unique.

Let x be an indeterminate over the real field. Then, for any $N = 1, 2, 3, \cdots$

$$\sum_{k=0}^{N-1} a_{2k} x^{2k} + \sum_{k=0}^{N-1} a_{2k+1} x^{2k+1} \equiv \sum_{l=0}^{N-1} (\alpha_l + x\beta_l)(1 + x^2)^l, \tag{2.4}$$

where the real numbers α_l and β_l are uniquely determined by

$$\alpha_l = \sum_{k=l}^{N-1}(-1)^{k-l}\binom{k}{l}a_{2k},$$

(2.5)

$$\beta_l = \sum_{k=l}^{N-1}(-1)^{k-l}\binom{k}{l}a_{2k+1},$$

$(l = 0, 1, \cdots, N-1)$. Thus, we have

THEOREM 2.1. *Every number* $\sum_{i=0}^{2N-1} a_i j_N{}^i$ *of* A_{2N} *may be expressed in a unique way in the form* $\sum_{l=0}^{N-1} (\alpha_l + j_N\beta_l)(1 + j_N{}^2)^l$.

B. In what follows, we shall make use of the fact that $j_N{}^2$ has an inverse. In order to obtain this inverse, we need the binomial identity

$$\sum_{k=q-1}^{N-1}\binom{k}{q-1} = \binom{N}{q}, \tag{2.6}$$

where $N = 1, 2, \cdots$ and $1 \leq q \leq N-1$. Using (2.6), we verify that

$$1/j_N{}^2 = -\sum_{k=0}^{N-1}(1 + j_N{}^2)^k. \tag{2.7}$$

C. Let the sequence of positive real numbers p_i be defined by the equations

$$\sum_{i+l=k} p_i p_l = 1, \qquad p_0 = 1, \tag{2.8}$$

$k = 0, 1, 2, \cdots$. It follows that $p_i = (-1)^i \binom{-\frac{1}{2}}{i}$, since $(\sum_{i=0}^{\infty} p_i x^i)^2 = 1/1 - x$ for $|x| < 1$, but the exact values of the p_i will not be needed. We define

$$i_N = j_N \sum_{i=0}^{N-1} p_i(1 + j_N{}^2)^i \tag{2.9}$$

for $N = 1, 2, \cdots$. Employing (2.7) and (2.9), we have

THEOREM 2.2. *The numbers* i_N *and* $-i_N$ *of the algebra* A_{2N} *satisfy the equation* $x^2 = -1$.

THEOREM 2.3. *Every number* $\sum_{k=0}^{N-1} (\alpha_k + j_N\beta_k)(1 + j_N{}^2)^k$ *of* A_{2N} *may be expresed in a unique way in the form* $\sum_{k=0}^{N-1} (\gamma_k + i_N\delta_k)(1 + j_N{}^2)^k$, *where* i_N *is given by* (2.9); *the real numbers* γ_k *and* δ_k *are given by*

$$(2.10) \qquad \begin{aligned} \gamma_k &= \alpha_k, \\ \sum_{i+l=k} \delta_i p_l &= \beta_k, \end{aligned}$$

($k = 0, 1, \cdots, N-1$), *and the p_l are given by* (2.8).

This may be seen at once by direct computation. The last result may be stated in another way. For $N = 1, 2, 3, \cdots$ let B_N denote the algebra of order N over the complex field whose elements are the polynomials

$$\sum_{i=0}^{N-1} b_i \omega_N{}^i,$$

where the b_i are complex numbers and the unit ω_N satisfies the equation

$$(\omega_N)^N = 0.$$

Equality is defined by postulating that

$$\sum_{i=0}^{N-1} b_i \omega_N{}^i = \sum_{i=0}^{N-1} c_i \omega_N{}^i,$$

if, and only if, $b_i = c_i$ for $i = 0, 1, \cdots, N-1$. Addition and multiplication are defined by the identities

$$(2.11) \qquad \begin{aligned} \sum_{i=0}^{N-1} b_i \omega_N{}^i + \sum_{i=0}^{N-1} c_i \omega_N{}^i &= \sum_{i=0}^{N-1} (b_i + c_i) \omega_N{}^i, \\ \left(\sum_{i=0}^{N-1} b_i \omega_N{}^i\right) \left(\sum_{i=0}^{N-1} c_i \omega_N{}^i\right) &= \sum_{i=0}^{N-1} \left(\sum_{l+k=i} b_l c_k\right) \omega_N{}^i. \end{aligned}$$

Theorem 2.3 states that for $N = 1, 2, 3, \cdots$ *the algebra A_{2N} of order $2N$ over the real field is isomorphic to the algebra B_N of order N over the complex field.* In this isomorphism, the number $\sum_{k=0}^{N-1} (\alpha_k + j_N \beta_k)(1 + j_N{}^2)^k$ of A_{2N} corresponds to the number $\sum_{k=0}^{N-1} (\gamma_k + i\delta_k) \omega_N{}^k$ of B_N, where the γ_k and δ_k are given by (2.10) and $i = j_1 = i_1$ is the ordinary complex unit.

D. Using the same notation as above, let

$$\sum_{k=0}^{N-1} a_{2k} j_N{}^{2k} + \sum_{k=0}^{N-1} a_{2k+1} j_N{}^{2k+1} = \sum_{k=0}^{N-1} (\alpha_k + j_N \beta_k)\ (1 + j_N{}^2)^k,$$

be a number of A_{2N}. Recalling (2.5) and (2.10), and keeping in mind that $p_0 = 1$ from (2.8), we have

$$\gamma_0 = \alpha_0 = \sum_{k=0}^{N-1} (-1)^k a_{2k}, \tag{2.12}$$
$$\delta_0 = \beta_0 = \sum_{k=0}^{N-1} (-1)^k a_{2k+1}.$$

THEOREM 2.4. *The number* $\sum_{i=0}^{2N-1} a_i j_N{}^i = \sum_{k=0}^{N-1} (\gamma_k + i_N \delta_k)(1 + j_N{}^2)^k$ *of* A_{2N} *has an inverse if, and only if, the ordinary complex number* $\gamma_0 + i\delta_0$ *does not vanish.*

Proof. In view of the isomorphism between A_{2N} and B_N, it suffices to show that the asserted condition is necessary and sufficient for the number $\sum_{i=0}^{N-1} b_i \omega_N{}^i = \sum_{k=0}^{N-1} (\gamma_k + i\delta_k) \omega_N{}^k$ of B_N to have an inverse.

Suppose that $b_0 = \gamma_0 + i\delta_0 = 1/c_0$ and define the complex numbers c_i by

$$b_0 c_0 = 1, \tag{2.13}$$
$$\sum_{l+k=i} b_l c_k = 0,$$

$(i = 1, 2, \cdots, N-1)$. Then, by (2.11), $\sum_{i=0}^{N-1} c_i \omega_N{}^i$ is the desired inverse. Conversely, if $\sum_{i=0}^{N-1} b_i \omega_N{}^i$ has the inverse $\sum_{k=0}^{N-1} c_i \omega_N{}^i$ then, from (2.11), equations (2.13) hold and c_0 is the inverse of b_0.

Taking (2.12) into account, it follows that the number $\sum_{i=0}^{2N-1} a_i j_N{}^i$ of A_{2N} has an inverse if, and only if,

$$\left[\sum_{k=0}^{N-1} (-1)^k a_{2k}\right]^2 + \left[\sum_{k=0}^{N-1} (-1)^k a_{2k+1}\right]^2 > 0. \tag{2.14}$$

THEOREM 2.5. *If a number* $\sum_{i=0}^{2N-1} a_i j_N{}^i$ *of* A_{2N} *which is the sum of even (odd) powers of* j_N *has an inverse, then its inverse is also the sum of even (odd) powers of* j_N.

Proof. Let

$$\sum_{k=0}^{N-1} a_{2k} j_N{}^{2k} + \sum_{k=0}^{N-1} a_{2k+1} j_N{}^{2k+1} = \sum_{k=0}^{N-1} (\alpha_k + j_N \beta_k)(1 + j_N{}^2)^k \tag{2.15}$$
$$= \sum_{k=0}^{N-1} (\gamma_k + i_N \delta_k)(1 + j_N{}^2)^k,$$

in A_{2N} and let

$$\sum_{i=0}^{N-1} b_i \omega_N{}^i = \sum_{k=0}^{N-1} (\gamma_k + i\delta_k) \omega_N{}^k, \tag{2.16}$$

be the image of $\sum_{i=0}^{2N-1} a_i j_N{}^i$ in B_N.

We note that all the $a_{2k} = 0$ in (2.15) if, and only if, all the b_i are purely imaginary in (2.16). Analogously, all the $a_{2k+1} = 0$ in (2.15) if, and only if, all the b_i are real in (2.16).

For, if all $a_{2k} = 0$ in (2.15), then all $\alpha_k = 0$ from (2.5), and all $\gamma_k = 0$ from (2.10). Hence, all the b_i of (2.16) are purely imaginary. Conversely, if all the b_i of (2.16) are purely imaginary, then all $\gamma_k = 0$ in (2.16), all $\alpha_k = 0$ from (2.10), and all $a_{2k} = 0$ from (2.5). Similarly for the a_{2k+1}.

Let the inverse of $\sum_{i=0}^{N-1} b_i \omega_N{}^i$ be denoted by $\sum_{i=0}^{N-1} c_i \omega_N{}^i$. To complete the proof, it is merely necessary to notice that in view of (2.11) and (2.13), if all the b_i are real (purely imaginary), then all the c_i are also real (purely imaginary), and to apply the above reasoning to $\sum_{i=0}^{N-1} c_i \omega_N{}^i$.

E. THEOREM 2.6. *If* $f = \sum_{i=0}^{2N-1} a_i j_N{}^i$ *and* $g = j_N f = \sum_{i=0}^{2N-1} b_i j_N{}^i$, *then*

$$\begin{aligned} b_{2k} &= a_{2k-1} - \binom{N}{k} a_{2N-1}, \\ b_{2k+1} &= a_{2k}, \end{aligned} \tag{2.17}$$

$(k = 0, \cdots, N-1)$, *and*

$$\begin{aligned} a_{2k} &= b_{2k+1}, \\ a_{2k+1} &= b_{2k+2} - \binom{N}{k+1} b_0, \end{aligned} \tag{2.18}$$

$(k = 0, \cdots, N-1)$.

(2.17) follows by direct computation, and (2.18) is obtained from it by observing that for $k = 0$ the first equation of (2.17) gives

$$b_0 = -a_{2N-1}.$$

We remark that if $g = j_N f$, then $f = g/j_N$, since j_N has an inverse, by (2.14).

Let $f = \sum_{i=0}^{2N-1} a_i j_N{}^i$ be a number of A_{2N}. We shall use the following notation:

$$\begin{aligned} Re[f] &= a_0, \\ Im[f] &= a_1. \end{aligned} \tag{2.19}$$

THEOREM 2.7. *Let* $f = \sum_{i=0}^{2N-1} a_i j_N{}^i$ *be a number of* A_{2N}, *and define*

$$K_{N,k} = -\sum_{l=k+1}^{N} \binom{N}{l} j_N^{2(l-k)}, \tag{2.20}$$

$$F_{N,k} = K_{N,k} f.$$

$(k = 0, \cdots, N-1)$. *Then*

$$Re[F_{N,k}] = a_{2k}, \qquad Im[F_{N,k}] = a_{2k+1}, \tag{2.21}$$

$(k = 0, \cdots, N-1)$.

Proof. From the definition (2.20), we have at once

$$K_{N,k} = j_N^2 K_{N,k+1} - j_N^2 \binom{N}{k+1}, \tag{2.22}$$

$(k = 0, 1, \cdots, N-2)$. The proof of the theorem now follows by an induction on k. First, $K_{N,k}$, for $k = N-1$, has the desired property (2.21), by (2.20) and (2.18). The induction is completed with the aid of these same two equations plus (2.22).

THEOREM 2.8. *$K_{N,k}$ has an inverse in A_{2N} which is the sum of even powers of j_N.*

Proof. The existence of the inverse follows immediately from (2.14), the definition of $K_{N,k}$, and the identity

$$-\sum_{l=k+1}^{N} \binom{N}{l} (-1)^{l-k} = \binom{N-1}{k},$$

$(k = 0, \cdots, N-1)$. That $1/K_{N,k}$ is a sum of even powers of j_N is seen from theorem 2.5.

F. In later sections, we shall employ the concept of the "absolute value," or norm, of a number of A_{2N}. Let $\alpha = \sum_{i=0}^{2N-1} a_i j_N^i$ be a number of A_{2N}. The absolute value of α, denoted by $\| \alpha \|$, is a non-negative number such that

$$\| \alpha \|^2 = \sum_{i=0}^{2N-1} a_i^2. \tag{2.23}$$

From this definition, we have the "triangle inequality"

$$| \| \alpha \| - \| \beta \| | \leqq \| \alpha + \beta \| \leqq \| \alpha \| + \| \beta \|. \tag{2.24}$$

If c is a real number, then

$$\| c\alpha \| = | c | \| \alpha \|. \tag{2.25}$$

There is a useful inequality which gives an upper bound for the absolute

value of the product of two numbers of A_{2N}. There exists a positive constant Γ_N such that [see (2), page 643]

$$\|\alpha\beta\| \leqq \Gamma_N \|\alpha\| \, \|\beta\|, \tag{2.26}$$

for any α and β of A_{2N}. Therefore

$$\|\alpha^n\| \leqq \Gamma_N^{n-1} \|\alpha\|^n. \tag{2.27}$$

G. Any hypercomplex number in the algebra A_{2N} may be written as a polynomial in j_N of degree at most $2N-1$. It is desirable to have a formula giving the powers of j_N greater than the $2N-1$-st as sums of powers of j_N ranging from 0 to $2N-1$. We have

$$\begin{aligned} j_N^{2\nu} &= (-1)^{N+\nu+1} \sum_{k=0}^{N-1} \binom{\nu}{k} \binom{\nu-1-k}{N-1-k} j_N^{2k}, \\ & \qquad\qquad\qquad (\nu = N, N+1, \cdots) \\ j_N^{2\nu+1} &= (-1)^{N+\nu+1} \sum_{k=0}^{N-1} \binom{\nu}{k} \binom{\nu-1-k}{N-1-k} j_N^{2k+1}. \end{aligned} \tag{2.28}$$

Only the first formula of (2.28) need be proved, since the second is a consequence of the first.

The proof is by induction, using the binomial identity

$$\binom{\nu}{k-1} \binom{\nu-k}{N-k} + \binom{\nu+1}{k} \binom{\nu-k}{N-k-1} = \binom{\nu}{N-1} \binom{N}{k}. \tag{2.29}$$

H. Expanding $(x+j_N y)^p$ by means of the binomial formula and replacing the powers of j_N above the $2N-1$-st by powers of j_N from the 0-th to the $2N-1$-st (using 2.28), we obtain

$$\begin{aligned} (x+j_N y)^p &= \sum_{k=0}^{N-1} \Big\{ \binom{p}{2k} y^{2k} x^{p-2k} + \sum_{\nu=N}^{[p/2]} \Big[(-1)^{N+\nu+1} \binom{p}{2\nu} \binom{\nu}{k} \binom{\nu-k-1}{N-k-1} y^{2\nu} x^{p-2\nu} \Big] \Big\} j_N^{2k} \\ &+ \sum_{k=0}^{N-1} \Big\{ \binom{p}{2k+1} y^{2k+1} x^{p-2k-1} + \sum_{\nu=N}^{[(p-1)/2]} \Big[(-1)^{N+\nu+1} \binom{p}{2\nu+1} \binom{\nu}{k} \binom{\nu-k-1}{N-k-1} \\ & \qquad y^{2\nu+1} x^{p-2\nu-1} \Big] \Big\} j_N^{2k+1}. \end{aligned} \tag{2.30}$$

For later reference, we shall write down (2.30) for even and odd p

$$\begin{aligned} (x+j_N y)^{2q} &= \sum_{k=0}^{N-1} \Big\{ \binom{2q}{2k} y^{2k} x^{2q-2k} + \sum_{\nu=N}^{q} \Big[(-1)^{N+\nu+1} \binom{2q}{2\nu} \binom{\nu}{k} \binom{\nu-k-1}{N-k-1} \\ & \qquad y^{2\nu} x^{2q-2\nu} \Big] \Big\} j_N^{2k} + \sum_{k=0}^{N-1} \Big\{ \binom{2q}{2k+1} y^{2k+1} x^{2q-2k-1} + \sum_{\nu=N}^{q-1} \Big[(-1)^{N+\nu+1} \\ & \qquad \binom{2q}{2\nu+1} \binom{\nu}{k} \binom{\nu-k-1}{N-k-1} y^{2\nu+1} x^{2q-2\nu-1} \Big] \Big\} j_N^{2k+1}; \end{aligned} \tag{2.31}$$

$$(2.32)\qquad (x+j_N y)^{2q+1} = \sum_{k=0}^{N-1}\{\binom{2q+1}{2k} y^{2k}x^{2q+1-2k} + \sum_{\nu=N}^{q}[(-1)^{N+\nu+1}\binom{2q-1}{2\nu} \binom{\nu}{k}\binom{\nu-k-1}{N-k-1} y^{2\nu}x^{2q+1-2\nu}]\}j_N{}^{2k} + \sum_{k=0}^{N-1}\{\binom{2q+1}{2k+1}y^{2k+1}x^{2q-2k} + \sum_{\nu=N}^{q}[(-1)^{N+\nu+1}\binom{2q+1}{2\nu+1}\binom{\nu}{k}\binom{\nu-k-1}{N-k-1} y^{2\nu+1}x^{2q-2\nu}]\}j_N{}^{2k+1}$$

To simplify the writing, we set

$$(2.33)\qquad \begin{aligned} (x+j_N y)^p &\equiv \sum_{i=0}^{2N-1}[\sum_{l=0}^{p} A_{ilp}{}^{(N)} y^l x^{p-l}] j_N{}^i, \\ j_N(x+j_N y)^p &\equiv \sum_{i=0}^{2N-1}[\sum_{l=0}^{p} B_{ilp}{}^{(N)} y^l x^{p-l}] j_N{}^i, \end{aligned}$$

where the $A_{ilp}{}^{(N)}$ are given explicitly by (2.30) and the $B_{ilp}{}^{(N)}$ may be obtained immediately using Theorem 2.6.

3. (Σ, N)-monogenic functions. Differentiation and integration.

A. For each positive integer N, we consider the system of equations $E(\Sigma, N)$

$$(3.1)\qquad \begin{aligned} \sigma_1(x)a_{2k,x} &= \tau_1(y)a_{2k+1,y}, \\ \sigma_2(x)a_{2k,y} &= \tau_2(y)[a_{2k-1,x} - a_{2N-1,x}\binom{N}{k}], \end{aligned}$$

$(k=0,\cdots,N-1)$, where $a_i = a_i(x,y)$ for $i=0,\cdots,2N-1$ and $a_{-1}\equiv 0$.

The system (3.1) may be written in a different way. For $k=0$, the second equation of (3.1) yields

$$\sigma_2(x)a_{0,y} = -\tau_2(y)a_{2N-1,x}.$$

With the aid of this last equation, (3.1) becomes

$$(3.2)\qquad \begin{aligned} \tau_2(y)a_{2k+1,x} &= \sigma_2(x)[a_{2k+2,y} - a_{0,y}\binom{N}{k+1}], \\ \tau_1(y)a_{2k+1,y} &= \sigma_1(x)a_{2k,x}, \end{aligned}$$

$(k=0,\cdots,N-1)$, where $a_{2N}\equiv 0$.

To simplify the writing, we shall use the condensed notation introduced in **1** and rewrite the last two equations in the form

$$(3.3)\qquad \begin{aligned} a_{2k,X} &= a_{2k+1,Y}, \\ a_{2k,Y^*} &= a_{2k-1,X^*} - a_{2N-1,X^*}\binom{N}{k}, \end{aligned}$$

$(k=0,\cdots,N-1)$, and

$$a_{2k+1,X^*}=a_{2k+2,Y^*}-a_{0,Y^*}\binom{N}{k+1}, \qquad (3.4)$$
$$a_{2k+1,Y}=a_{2k,X},$$

$(k=0,\cdots,N-1)$.

By analogy with the theory of functions of a complex variable, we lay down the following definitions:

DEFINITION 3.1. *A hypercomplex valued function* $f(z)=\sum_{i=0}^{2N-1}a_ij_N{}^i$, *where* $a_i=a_i(x,y)$ *for* $i=0,\cdots,2N-1$, *is said to be* (Σ,N)*-monogenic at a point* $z=(x,y)$ *if there exists a neighborhood* D *of* z *such that the functions* a_i *belong to class* $C^{(1)}$ *in* D *and satisfy the system* $E(\Sigma,N)$ *in* D.

We shall say that $f(z)=\sum_{i=0}^{2N-1}a_ij_N{}^i$ is (Σ,N)-monogenic in the domain D if $f(z)$ is (Σ,N)-monogenic at all points of D. With this understanding, we have the

DEFINITION 3.2. *If the function* $f(z)=\sum_{i=0}^{2N-1}a_ij_N{}^i$ *is* (Σ,N)*-monogenic in a domain* D *and all the* a_i *belong to class* $C^{(2)}$ *in* D, *then the function* $f^{[1]}(z)=\sum_{i=0}^{2N-1}a_i{}^{[1]}j_N{}^i$, *where*

$$a_{2k}{}^{[1]}=a_{2k,X}=a_{2k+1,Y}, \qquad (3.5)$$
$$a_{2k+1}{}^{[1]}=a_{2k+1,X^*}=a_{2k+2,Y^*}-a_{0,Y^*}\binom{N}{k+1},$$

$(k=0,\cdots,N-1)$, *is said to be the* (Σ,N)*-derivative of* $f(z)$ *and is denoted by* $d_{(\Sigma,N)}f/d_{(\Sigma,N)}z$ *or more simply by* $d_\Sigma f/d_\Sigma z$ *when no confusion as to the value of* N *might arise.*

From (3.5) and Theorem 2.6, we may write

$$\begin{aligned}f^{[1]}(z)&=\sum_{k=0}^{N-1}a_{2k,X}j_N{}^{2k}+\sum_{k=0}^{N-1}a_{2k+1,X^*}j_N{}^{2k+1}\\&=\sum_{k=0}^{N-1}a_{2k+1,Y}j_N{}^{2k}+\sum_{k=0}^{N-1}\left[a_{2k+2,Y^*}-a_{0,Y^*}\binom{N}{k+1}\right]j_N{}^{2k+1}\\&=(1/j_N)\left[\sum_{k=0}^{N-1}a_{2k,Y^*}j_N{}^{2k}+\sum_{k=0}^{N-1}a_{2k+1,Y}j_N{}^{2k+1}\right].\end{aligned}$$

The following remarks are immediate consequences of the last two definitions: Any hypercomplex constant $\sum_{i=0}^{2N-1}b_ij_N{}^i$ is (Σ,N)-monogenic, and

conversely any (Σ, N)-monogenic function whose (Σ, N)-derivative vanishes identically is a constant. Furthermore, if f and g are (Σ, N)-monogenic, and α and β are *real* constants, then $\alpha f + \beta g$ is (Σ, N)-monogenic, and

$$[\alpha f + \beta g]^{[1]} = \alpha f^{[1]} + \beta g^{[1]}.$$

We also have

THEOREM 3.1. *If the function* $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ *is* (Σ, N)*-monogenic and possesses a* (Σ, N)*-derivative* $f^{[1]}(z)$ *in a domain* D, *then* $f^{[1]}(z)$ *is* (Σ', N)*-monogenic in* D.

Proof. From (3.5), it follows that

$$a^{[1]}{}_{2k,X^*} = a_{2k+1,YX^*}, \qquad a^{[1]}{}_{2k+1,Y} = a_{2k+1,X^*Y},$$
$$a^{[1]}{}_{2k,Y^*} = a_{2k,XY^*}, \qquad a^{[1]}{}_{2k+1,X} = a_{2k+2,Y^*X} - a_{0Y^*X}\binom{N}{k+1},$$

$(k = 0, \cdots, N-1)$, and since

$$a^{[1]}{}_{2N-1,X} = -a_{0,Y^*X},$$

we obtain

$$\begin{aligned} a^{[1]}{}_{2k,X^*} &= a^{[1]}{}_{2k+1,Y}, \\ a^{[1]}{}_{2k,Y^*} &= a^{[1]}{}_{2k-1,X} - a^{[1]}{}_{2N-1,X}\binom{N}{k}, \end{aligned} \tag{3.6}$$

$(k = 0, \cdots, N-1)$, which is $E(\Sigma', N)$, by (1.14).

Next, we proceed to the definition of higher (Σ, N)-derivatives of a (Σ, N)-monogenic function. If the function $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ is (Σ, N)-monogenic in a domain D and all the a_i belong to class $C^{(m)}$, $m \geq 2$, then the higher (Σ, N)-derivatives $f^{[n]}(z) = \sum_{i=0}^{2N-1} a_i{}^{[n]} j_N{}^i$ are defined by

$$f^{[0]}(z) = f(z),$$

$$f^{[n]}(z) = \begin{cases} \dfrac{d_{(\Sigma,N)}\, f^{[n-1]}(z)}{d_{(\Sigma,N)}\, z} & \text{if } n \text{ is odd}, \\[2ex] \dfrac{d_{(\Sigma',N)}\, f^{[n-1]}(z)}{d_{(\Sigma',N)}\, z} & \text{if } n \text{ is even}, \end{cases} \tag{3.7}$$

$(n = 1, 2, \cdots, m-1)$. Obviously, $f^{[2n]}(z)$ is (Σ, N)-monogenic and $f^{[2n+1]}(z)$ is (Σ', N)-monogenic.

B. The (Σ, N)-integral of a continuous function $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ over a rectifiable curve C is defined by

$$(3.8) \qquad \int_C f(z)\, d_{(\Sigma,N)} z = \sum_{i=0}^{2N-1} A_i j_N{}^i,$$

where

$$A_{2k} = \int_C a_{2k} dX^* + [a_{2k-1} - a_{2N-1} \binom{N}{k}] dY^*$$

$$(3.9) \qquad = \int_C \sigma_2 a_{2k} dx + \tau_2 [a_{2k-1} - a_{2N-1} \binom{N}{k}] dy,$$

$$A_{2k+1} = \int_C a_{2k+1} dX + a_{2k} dY$$

$$= \int_C (1/\sigma_1) a_{2k+1} dx + (1/\tau_1) a_{2k} dy,$$

$(k = 0, \cdots, N-1)$.

Introducing the ordinary complex-valued functions

$$(3.10) \qquad \begin{aligned} h_{2k}(z) &= \sigma_2 a_{2k} - i\tau_2 [a_{2k-1} - a_{2N-1} \binom{N}{k}], \\ h_{2k+1}(z) &= (1/\tau_1) a_{2k} + i(1/\sigma_1) a_{2k+1}, \end{aligned}$$

$(k = 0, \cdots, N-1)$, and recalling the expression for $\int_C g(z)dz$, where $dz = dx + idy$, (the ordinary complex integral over C of the continuous function $g(z) = s + it$), equation (3.8) may be rewritten in the form

$$(3.11) \qquad \int_C f(z)\, d_{(\Sigma,N)} z = \sum_{k=0}^{N-1} \{Re[\int_C h_{2k}(z)\, dz]\} j_N{}^{2k}$$

$$+ \sum_{k=0}^{N-1} \{Im[\int_C h_{2k+1}(z)\, dz]\} j_N{}^{2k+1}.$$

From (3.8) and (3.9), it follows that if $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ and $g(z) = \sum_{i=0}^{2N-1} b_i j_N{}^i$ are continuous on C and α and β are *real* constants, then

$$\int_C (\alpha f + \beta g)\, d_{(\Sigma,N)} z = \alpha \int_C f d_{(\Sigma,N)} z + \beta \int_C g d_{(\Sigma,N)} z.$$

Also, over C

$$\int_a^b f d_{(\Sigma,N)} z + \int_b^c f d_{(\Sigma,N)} z = \int_a^c f d_{(\Sigma,N)} z.$$

Let l be the length of C; A, a constant such that for all points of C, $|x| \leqq A$, $|y| \leqq A$, and $m_\Sigma(A)$ the non-decreasing function defined by $m_\Sigma(A) = \max_{i=1,2} \max_{l=\pm 1} \max_{|t| \leqq A} \{\sigma_i(t)^l, \tau_i(t)^l\}$. The following inequality holds

$$\| \int_C fd_{(\Sigma,N)}z \| \leq (2^N + 2N) m_\Sigma (A) \cdot \max_C \| f \| \cdot l. \tag{3.12}$$

For, from (3.11)

$$\| \int_C f(z) d_{(\Sigma,N)} z \| \leq (\sum_{k=0}^{N-1} \max_C | h_{2k} | + \sum_{k=0}^{N-1} \max_C | h_{2k+1} |) l,$$

while from (3.10), letting $M = \max_C \| f \|$ and $m = m_\Sigma(A)$

$$| h_{2k+1} | \leq 2mM,$$

$(k = 0, \cdots, N-1)$, and

$$| h_0 | \leq 2mM,$$
$$| h_{2k} | \leq mM(2 + \binom{N}{k}).$$

THEOREM 3.2. *If* $f(z) = \sum_{l=0}^{2N-1} a_i j_N{}^i$ *is* (Σ, N)*-monogenic in a closed simply connected domain* D *bounded by a rectifiable curve* C *then*

$$\int_C f(z) d_{(\Sigma,N)} z = 0. \tag{3.13}$$

Proof. This follows immediately upon observing that the line integrals in (3.9) are all zero in view of (3.1) or (3.3).

In view of the independence of the path asserted by Theorem 3.2, we may consider the function

$$F(z) = \int_{z_0}^{z} fd_{(\Sigma,N)}z = \sum_{i=0}^{2N-1} A_i j_N{}^i,$$

where $f = \sum_{i=0}^{2N-1} a_i j_N{}^i$ is (Σ, N)-monogenic in D. It is easily shown that

THEOREM 3.3. *The* (Σ, N)*-integral of a* (Σ, N)*-monogenic function is* (Σ', N)*-monogenic.*

For, from (3.9)

$$\begin{array}{ll} A_{2k,X^*} = a_{2k}, & A_{2k+1,Y} = a_{2k}, \\ A_{2k,Y^*} = a_{2k-1} - a_{2N-1} \binom{N}{k}, & A_{2k+1,X} = a_{2k+1}, \end{array} \tag{3.14}$$

$(k = 0, \cdots, N-1)$, and since

$$A_{2N-1,X} = a_{2N-1},$$

we obtain $E(\Sigma', N)$ from (3.14).

From (3.14) and Theorem 3.3, by a simple induction, we may prove the following corollary

COROLLARY 3.3. *Let* $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ *be* (Σ, N)*-monogenic in a simply connected domain* D *and define*

$$F_0(z) = f(z),$$

$$(3.15) \qquad F_n(z) = \begin{cases} \displaystyle\int_a^z F_{n-1}(\xi)\, d_{(\Sigma',N}\, \xi & \text{if } n \text{ is odd,} \\ \displaystyle\int_a^z F_{n-1}(\xi)\, d_{(\Sigma',N}\, \xi & \text{if } n \text{ is even,} \end{cases}$$

for $n \geqq 1$, *where* a *is a fixed point of* D. *Then* $F_n(z) = \sum_{i=0}^{2N-1} A_i{}^{(n)} j_N{}^i$ *is* (Σ, N)*-monogenic for* n *even and* (Σ', N)*-monogenic for* n *odd. Furthermore, the functions* $A_i{}^{(n)}$ *belong to class* $C^{(n)}$ *and for* $q = 0, 1, \cdots$

$$(3.16) \qquad \begin{aligned} &\frac{\partial^{2q} A_{2k}{}^{(2q)}}{(\partial X^* \partial X)^q} = a_{2k}, & &\frac{\partial^{2q} A_{2k+1}{}^{(2q)}}{(\partial X \partial X^*)^q} = a_{2k+1}, \\ &\frac{\partial^{2q+1} A_{2k}{}^{(2q+1)}}{\partial X^* (\partial X \partial X^*)^q} = a_{2k}, & &\frac{\partial^{2q+1} A_{2k+1}{}^{(2q+1)}}{\partial X (\partial X^* \partial X)^q} = a_{2k+1}, \end{aligned}$$

$(k = 0, \cdots, N-1)$.

4. Contraction. In the preceding section, we defined (Σ, N)- and (Σ', N)-differentiation and integration. By repeated application of these processes, we can obtain new solutions of the system $E(\Sigma, N)$ from any given solution of this system. Here, we shall be concerned with a process which enables us to find solutions of the systems $E(\Sigma, M)$, $1 \leqq M \leqq N-1$, given a solution of the system $E(\Sigma, N)$.

LEMMA 4.1. *If* $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ *is* (Σ, N)*-monogenic in a domain* D *then* $\phi(z) = \sum_{i=0}^{2N-3} \alpha_i j_{N-1}{}^i$, *where*

$$(4.1) \qquad \begin{aligned} \alpha_{2k} &= a_{2k} - \binom{N-1}{k} a_{2N-2}, \\ \alpha_{2k+1} &= a_{2k+1} - \binom{N-1}{k} a_{2N-1}, \end{aligned}$$

$(k = 0, \cdots, N-2)$, *is* $(\Sigma, N-1)$*-monogenic in* D.

Proof. This is easily verified using (4.1) and the binomial identity

$$\binom{N-1}{k-1}+\binom{N-1}{k}=\binom{N}{k}.$$

THEOREM 4.1. *If the function* $f_N(z)=\sum_{k=0}^{N-1} a_{2k}j_N{}^{2k}+\sum_{k=0}^{N-1} a_{2k+1}j_N{}^{2k+1}$ *is* (Σ, N)*-monogenic in a domain* D *then for each* $i=0,\cdots,N-1$ *the function*

$$\begin{aligned} f_{N-i}(z) &= \sum_{k=0}^{N-1} a_{2k}j_{N-i}{}^{2k}+\sum_{k=0}^{N-1} a_{2k+1}j_{N-i}{}^{2k+1} \\ &= \sum_{k=0}^{N-i-1}\Big[a_{2k}+\sum_{\nu=N-i}^{N-1}\Big\{(-1)^{N-i+\nu+1}\binom{\nu}{k}\binom{\nu-k-1}{N-i-k-1}a_{2\nu}\Big\}\Big]j_{N-i}{}^{2k} \\ &\quad+\sum_{k=0}^{N-i-1}\Big[a_{2k+1}+\sum_{\nu=N-i}^{N-1}\Big\{(-1)^{N-i+\nu+1}\binom{\nu}{k}\binom{\nu-k-1}{N-i-k-1}a_{2\nu+1}\Big\}\Big]j_{N-i}{}^{2k+1}, \end{aligned} \tag{4.2}$$

is $(\Sigma, N-i)$*-monogenic in* D.

A sketch of the proof is as follows. First, using (2.28), we verify that the formal substitution of j_{N-i} for j_N in $f_N(z)$ yields the expression on the extreme right of (4.2). To prove that $f_{N-i}(z)$ is $(\Sigma, N-i)$-monogenic, we proceed by induction, employing Lemma 4.1 and the binomial identity

$$\binom{\nu}{k}\binom{\nu-k-1}{N-i-k-1}+\binom{\nu}{k}\binom{\nu-k-1}{N-i-k-2} = \binom{N-i-1}{k}\binom{\nu}{N-i-1}.$$

In particular, we observe that taking $i=N-1$ shows that

$$f_1(z)=\sum_{k=0}^{N-1}(-1)^k a_{2k}+j_1\sum_{k=0}^{N-1}(-1)^k a_{2k+1},$$

is always $(\Sigma, 1)$-monogenic.

5. The partial differential equations of the 2Nth order. The main purpose of this section is to show that if $f(z)=\sum_{i=0}^{2N-1} a_i j_N{}^i$ is (Σ, N)-monogenic in a domain D, then the real functions a_{2k} $(k=0,\cdots,N-1)$ satisfy in D the partial differential equation of the $2N$-th order

$$\left(\frac{\partial^2}{\partial X^*\partial X}+\frac{\partial^2}{\partial Y\partial Y^*}\right)^N u=0,$$

and the real functions a_{2k+1} $(k=0,\cdots,N-1)$ satisfy in D the partial differential equation of the $2N$-th order

$$\left(\frac{\partial^2}{\partial X \partial X^*} + \frac{\partial^2}{\partial Y^* \partial Y}\right)^N u = 0.$$

In addition, it will be shown that all the theorems of Section **3** hold under the sole assumption that the functions a_i belong to class $C^{(1)}$.

LEMMA 5.1. *If* $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ *is* (Σ, N)*-monogenic in a domain* D *and the* a_i *belong to class* $C^{(3)}$ *in* D*, then the function* $g(z) = \sum_{i=0}^{2N-1} b_i j_N{}^i$*, where*

$$b_{2k} = a_{2k+1, X^*Y} = (\tau_1/\sigma_2) a_{2k+1, xy},$$

$$b_{2k+1} = a_{2k+2, XY^*} - a_{0, XY^*}\left(\frac{N}{k+1}\right) = (\sigma_1/\tau_2)\left[a_{2k+2, xy} - a_{0, xy}\left(\frac{N}{k+1}\right)\right],$$

$(k = 0, \cdots, N-1)$*, is also* (Σ, N)*-monogenic in* D.

Proof. It is readily verified, from (3.7) and (3.5), that $g(z) = f^{[2]}(z)$. That $g(z)$ is (Σ, N)-monogenic follows from the remark immediately after equation (3.7).

THEOREM 5.1. *If* $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ *is* (Σ, N)*-monogenic in a domain* D *and the* a_i *belong to class* $C^{(2N)}$ *in* D*, then*

$$L^N a_{2k} = 0, \quad M^N a_{2k+1} = 0, \tag{5.1}$$

$(k = 0, \cdots, N-1)$*, in* D*, where* L *and* M *are the linear differential operators defined by*

$$\begin{aligned} Lu &\equiv (\tau_1/\sigma_2)[(\sigma_1 u_x/\tau_1)_x + (\sigma_2 u_y/\tau_2)_y] = U_{XX^*} + U_{Y^*Y}, \\ Mu &\equiv (\sigma_1/\tau_2)[(\tau_2 u_x/\sigma)_x + (\tau_1 u_y/\sigma_1)_y] = U_{X^*X} + U_{YY^*}, \end{aligned} \tag{5.2}$$

Proof. We proceed by induction. $E(\Sigma, 1)$ is the system of equations considered by Bers and Gelbart (1)

$$\begin{aligned} \sigma_1 a_{0,x} &= \tau_1 a_{1,y} \\ \sigma_2 a_{0,y} &= -\tau_2 a_{1,x}, \end{aligned} \quad \text{or} \quad \begin{aligned} a_{0,X} &= a_{1,Y}, \\ a_{0,Y^*} &= -a_{1,X^*}, \end{aligned}$$

Assuming that a_1 and a_0 belong to class $C^{(2)}$, and eliminating a_1 and a_0 in turn, we obtain

$$La_0 = 0,$$
$$Ma_1 = 0,$$

and the conclusion of the theorem has been verified for $N = 1$.

Suppose now that the conclusion is true for $N-1\geqq 1$ and let $f(z)=\sum_{i=0}^{2N-1} a_i j_N{}^i$ be (Σ, N)-monogenic, where the a_i belong to class $C^{(2N)}$. We have $E(\Sigma, N)$

$$
\begin{aligned}
&a_{2k,X}=a_{2k+1,Y},\\
&a_{2k,Y^*}=a_{2k-1,X^*}-a_{2N-1,X^*}\binom{N}{k},
\end{aligned}
\tag{5.3}
$$

$(k=0,\cdots,N-1)$. By Lemma 4.1, the function $\phi(z)=\sum_{i=0}^{2N-3}\alpha_i j_{N-1}{}^i$ is $(\Sigma, N-1)$-monogenic, where

$$
\begin{aligned}
\alpha_{2k}&=a_{2k}-\binom{N-1}{k}a_{2N-2},\\
\alpha_{2k+1}&=a_{2k+1}-\binom{N-1}{k}a_{2N-1},
\end{aligned}
\tag{5.4}
$$

$(k=0,\cdots,N-2)$. Consequently, by the induction hypothesis, since the α_i belong to the class $C^{(2N)}$ in D

$$
\begin{aligned}
L^{N-1}\alpha_{2k}&=0,\\
M^{N-1}\alpha_{2k+1}&=0,
\end{aligned}
\tag{5.5}
$$

$(k=0,\cdots,N-2)$. Since L and M are linear operators, (5.4) and (5.5) imply

$$
\begin{aligned}
L^{N-1}a_{2k}&=\binom{N-1}{k}L^{N-1}a_{2N-2},\\
M^{N-1}a_{2k+1}&=\binom{N-1}{k}M^{N-1}a_{2N-1},
\end{aligned}
\tag{5.6}
$$

$(k=0,\cdots,N-2)$. It is obvious then that to verify (5.1) it suffices to show that $L^N a_{2N-2}$ and $M^N a_{2N-1}$ are both zero.

From (5.3), for $k=0$,

$$
\begin{aligned}
&a_{0,X}=a_{1,Y},\\
&a_{0,Y^*}=-a_{2N-1,X^*},
\end{aligned}
\tag{5.7}
$$

Using the definition of L, (5.2) and (5.4), equation (5.7) yields

$$La_0=\alpha_{1,X^*Y}, \tag{5.8}$$

since

$$
\begin{aligned}
La_0&=a_{0,XX^*}+a_{0,Y^*Y}\\
&=a_{1,YX^*}-a_{2N-1,X^*Y}=\left[a_1-\binom{N-1}{0}a_{2N-1}\right]_{X^*Y}.
\end{aligned}
$$

From (3.4) for $k=0$

$$
\begin{aligned}
&a_{1,X^*} = a_{2,Y^*} - a_{0,Y^*}(N-1), \\
&a_{1,Y} = a_{0,X}.
\end{aligned}
\tag{5.9}
$$

Using the definition of M, (5.2), and (5.4), equation (5.9) yields

$$
Ma_1 = \alpha_{2,XY^*} - \binom{N-1}{1}\alpha_{0,XY^*},
\tag{5.10}
$$

since

$$
\begin{aligned}
Ma_1 &= a_{1,X^*X} + a_{1,YY^*} \\
&= a_{2,Y^*X} - a_{0,Y^*X}(N-1) + a_{0,XY^*} \\
&= \left[a_2 - \binom{N-1}{1} a_0\right]_{XY^*},
\end{aligned}
$$

and from (5.4)

$$
\alpha_2 - \binom{N-1}{1}\alpha_0 = \left[a_2 - \binom{N-1}{1} a_{2N-2}\right] - \binom{N-1}{1}\left[a_0 - \binom{N-1}{0} a_{2N-2}\right] = a_2 - \binom{N-1}{1} a_0.
$$

On the other hand, since $N \geqq 2$, we may apply Lemma 5.1 to $\phi(z)$ to prove that the function $\phi^{[2]}(z) = \sum_{i=0}^{2N-3} \alpha_i^{[2]} j_{N-1}{}^i$ is $(\Sigma, N-1)$-monogenic, where

$$
\begin{aligned}
&\alpha_{2k}{}^{[2]} = \alpha_{2k+1,X^*Y}, \\
&\alpha_{2k+1}{}^{[2]} = \alpha_{2k+2,XY^*} - \alpha_{0,XY^*}\binom{N-1}{k+1},
\end{aligned}
\tag{5.11}
$$

$(k=0,\cdots,N-2)$. Consequently, by the induction hypothesis, inasmuch as the α_i belong to class $C^{(2N-2)}$, we have

$$
\begin{aligned}
&L^{N-1}\alpha_{2k}{}^{[2]} = 0, \\
&M^{N-1}\alpha_{2k+1}{}^{[2]} = 0,
\end{aligned}
\tag{5.12}
$$

$(k=0,\cdots,N-2)$.

But a comparison of (5.8), and (5.10) with (5.11) gives

$$
\begin{aligned}
&La_0 = \alpha_0{}^{[2]}, \\
&Ma_1 = \alpha_1{}^{[2]},
\end{aligned}
\tag{5.13}
$$

and, in view of (5.12), (5.13) yields

$$
\begin{aligned}
&L^N a_0 = 0, \\
&M^N a_1 = 0.
\end{aligned}
\tag{5.14}
$$

The desired conclusion, (5.1), follows from (5.6) and (5.14).

It is now possible to remove most of the differentiability assumptions made on the a_i in the hypotheses of several theorems. It will be recalled that $\sigma_1, \sigma_2, \tau_1$, and τ_2 are positive analytic functions of their respective real variables. Hence, the operators L and M are of elliptic type with analytic coefficients and any function u of class $C^{(2N)}$ satisfying either $L^N u = 0$ or $M^N u = 0$ for a fixed $N = 1, 2, \cdots$, is necessarily analytic. Let $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ be (Σ, N)-monogenic in a simply connected domain D and consider the function

$$F_{2N}(z) = \sum_{i=0}^{2N-1} A_i{}^{(2N)} j_N{}^i,$$

(the functions $F_n(z)$ were defined in (3.15)). From Corollary 3.3, it follows that $F_{2N}(z)$ is (Σ, N)-monogenic and that the $A_i{}^{(2N)}$ belong to class $C^{(2N)}$. Hence, from Theorem 5.1

$$L^N A_{2k}{}^{(2N)} = 0,$$
$$M^N A_{2k+1}{}^{(2N)} = 0,$$

$(k = 0, \cdots, N-1)$, and consequently the $A_i{}^{(2N)}$ are analytic. But from (3.16)

$$\frac{\partial^{2N} A_{2k}{}^{(2N)}}{(\partial X^* \partial X)^N} = a_{2k}, \qquad \frac{\partial^{2N} A_{2k+1}{}^{(2N)}}{(\partial X \partial X^*)^N} = a_{2k+1},$$

$(k = 0, \cdots, N-1)$, which shows that $a_i(x, y)$ is a linear combination, with analytic coefficients, of $A_i{}^{(2N)}(x, y)$ and its partial derivatives with respect to x up to the $2N$-th order. Thus the functions $a_i(x, y)$ are analytic functions of x and y and we have the

THEOREM 5.2. *If $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ is (Σ, N)-monogenic in a domain D, then the functions $a_i(x, y)$ are analytic functions of x and y in D.*

Using this result, we see that unlimited successive (Σ, N)- and (Σ', N)-differentiations of a (Σ, N)-monogenic function are possible.

Theorem 5.1 may now be restated in the form

THEOREM 5.3. *If $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ is (Σ, N)-monogenic in a domain D, then*

$$L^N a_{2k} = 0,$$
$$M^N a_{2k+1} = 0,$$

$(k = 0, \cdots, N-1)$, in D.

We remark that (Σ, N)-integration and (Σ', N)-differentiation, as well as (Σ, N)-differentiation and (Σ', N)-integration, are inverse processes. That is, if $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ is (Σ, N)-monogenic, then

$$\frac{d_{(\Sigma',N)}}{d_{(\Sigma',N)} z} \int_{z_0}^{z} f(\zeta)\, d_{(\Sigma,N)}\, \zeta = f(z), \tag{5.15}$$

and

$$\int_{z_0}^{z} f^{[1]}(\zeta)\, d_{(\Sigma',N)}\, \zeta = f(z) - f(z_0), \tag{5.16}$$

Next, we prove the analogue of Morera's theorem.

THEOREM 5.4. *If $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ is continuous in a simply connected domain D and for every simple closed rectifiable curve C in D, $\int_C f(z)\, d_{(\Sigma,N)}\, z = 0$, then $f(z)$ is (Σ, N)-monogenic in D.*

Proof. Consider the function

$$F(z) = \int_{z_0}^{z} f d_{(\Sigma,N)}\, z = \sum_{i=0}^{2N-1} A_i j_N{}^i.$$

By (3.14), as in the proof of Theorem 3.3, we see that $f(z)$ is (Σ', N)-monogenic. By (3.14) and (3.5), $\dfrac{d_{(\Sigma',N)} F}{d_{(\Sigma',N)}\, z} = f(z)$.

THEOREM 5.5. *If a series of (Σ, N)-monogenic functions $\sum_{n=0}^{\infty} fn(z)$, where $f_n(z) = \sum_{n=0}^{\infty} a_{ni} j_N{}^i$, converges uniformly to $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ in a domain D, then $f(z)$ is (Σ, N) monogenic in D.*

6. The associated matrices. The system $E(\Sigma, N)$ has

$$\Sigma = \begin{Vmatrix} \sigma_1 & \tau_1 \\ \sigma_2 & \tau_2 \end{Vmatrix},$$

as the matrix of its coefficients. Together with $E(\Sigma, N)$, we shall be led to consider systems whose coefficient matrices are

$$\Sigma' = \begin{Vmatrix} 1/\sigma_2 & \tau_1 \\ 1/\sigma_1 & \tau_2 \end{Vmatrix}, \quad \Sigma_{\prime} = \begin{Vmatrix} \sigma_1 & 1/\tau_2 \\ \sigma_2 & 1/\tau_1 \end{Vmatrix}, \quad \Sigma_{\prime}{}' = \begin{Vmatrix} 1/\sigma_2 & 1/\tau_2 \\ 1/\sigma_1 & 1/\tau_1 \end{Vmatrix}. \tag{6.1}$$

Obviously,

$$\Sigma'' \equiv \Sigma_{\prime\prime} \equiv \Sigma_{\prime\prime}{}'' \equiv \Sigma.$$

It has already been shown that the (Σ, N)-integral and the (Σ, N)-derivative of a (Σ, N)-monogenic function are both (Σ', N)-monogenic functions. We now prove two theorems concerning the relations between (Σ, N)-, (Σ', N)-, $(\Sigma_{,}, N)$-, and $(\Sigma_{,}', N)$-monogenic functions.

THEOREM 6.1. *If* $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ *is* (Σ, N)*-monogenic, then* $g(z) = j_N f(z)$ *is* $(\Sigma_{,}', N)$*-monogenic.*

Letting $g(z) = \sum_{i=0}^{2N-1} b_i j_N{}^i$ and using Theorem 2.6, the result is immediate.

COROLLARY 6.1. *If* $f(z)$ *is* (Σ, N)*-monogenic, then* $j_N{}^{2k} f(z)$ *is* (Σ, N)*-monogenic and* $j_N{}^{2k+1} f(z)$ *is* $(\Sigma_{,}', N)$*-monogenic, where* k *is any integer.*

The proof follows for positive k by repeated application of Theorem 6.1 and then for negative k by using Theorem 2.5.

THEOREM 6.2. *If* $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ *is* (Σ, N)*-monogenic, then*

$$\frac{d_{(\Sigma_{,}', N)} [j_N f]}{d_{(\Sigma_{,}', N)} z} = j_N \frac{d_{(\Sigma, N)} f}{d_{(\Sigma, N)} z}, \tag{6.3}$$

and

$$\int_{z_0}^{z} j_N f d_{(\Sigma_{,}', N)} z = j_N \int_{z_0}^{z} f d_{(\Sigma, N)} z. \tag{6.4}$$

Proof. (6.3) is a consequence of the definition of the (Σ, N)-derivative equation (3.5), and Theorem 6.1. As regards (6.4), both sides of the proposed equality are $(\Sigma_{,}, N)$-monogenic functions. In addition, both functions vanish at z_0 and, by (6.3), the $(\Sigma_{,}, N)$-derivative of their difference vanishes identically.

COROLLARY 6.2. *If* $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ *is* (Σ, N)*-monogenic, then*

$$\frac{d_{(\Sigma, N)} [j_N{}^{2k} f]}{d_{(\Sigma, N)} z} = j_N{}^{2k} \frac{d_{(\Sigma, N)} f}{d_{(\Sigma, N)} z}, \tag{6.5}$$

$$\int_{z_0}^{z} j_N{}^{2k} f d_{(\Sigma, N)} z = j_N{}^{2k} \int_{z_0}^{z} f d_{(\Sigma, N)} z, \tag{6.6}$$

$$\frac{d_{(\Sigma_{,}', N)} [j_N{}^{2k+1} f]}{d_{(\Sigma_{,}', N)} z} = j_N{}^{2k+1} \frac{d_{(\Sigma, N)} f}{d_{(\Sigma, N)} z}, \tag{6.7}$$

$$\int_{z_0}^{z} j_N{}^{2k+1} f d_{(\Sigma_{,}', N)} z = j_N{}^{2k+1} \int_{z_0}^{z} f d_{(\Sigma, N)} z, \tag{6.8}$$

where k *is any integer.*

7. The formal powers.

A. Any hypercomplex constant is a (Σ, N)-monogenic function; hence, we may construct (Σ, N)-monogenic functions by successive (Σ, N)- and (Σ', N)-integrations of a constant. Accordingly, we define [see Bers and Gelbart (1), Section 4]

$$
\begin{aligned}
&a \cdot Z^{(0)}(N; z_0; z) = a \cdot \tilde{Z}^{(0)}(N; z_0; z) = a, \\
(7.1)\qquad &a \cdot Z^{(n)}(N; z_0; z) = n \int_{z_0}^{z} (a \cdot \tilde{Z}^{(n-1)}(N; z_0; z))\, d_{(\Sigma', N)} z, \\
&a \cdot \tilde{Z}^{(n)}(N; z_0; z) = n \int_{z_0}^{z} (a \cdot Z^{(n-1)}(N; z_0 z))\, d_{(\Sigma, N)} z,
\end{aligned}
$$

for $n \geqq 1$, where $a = \sum_{i=0}^{2N-1} \alpha_i j_N{}^i$ is a hypercomplex constant and the point z_0 is fixed. In this formula $a \cdot Z$ is *one* symbol and does not denote ordinary multiplication of a by Z, since no meaning has been attached to Z by itself as yet. In addition, we have

$$(7.2)\qquad a \cdot Z^{(n)}(N; z_0; z_0) = a \cdot \tilde{Z}^{(n)}(N; z_0; z_0) = 0$$

and

$$
\begin{aligned}
&\frac{d_{(\Sigma, N)}(a \cdot Z^{(n)}(N; z_0; z))}{d_{(\Sigma, N)} z} = na \cdot \tilde{Z}^{(n-1)}(N; z_0; z), \\
(7.3)\qquad &\frac{d_{(\Sigma', N)}(a \cdot \tilde{Z}^{(n)}(N; z_0; z))}{d_{(\Sigma', N)} z} = na \cdot Z^{(n-1)}(N; z_0; z).
\end{aligned}
$$

In this section, we shall keep N fixed. To simplify the writing, we set

$$1 \cdot Z^{(n)}(N; z_0; z) \equiv Z^{(n)}(z_0; z), \qquad a \cdot Z^{(n)}(N; z_0; z) \equiv a \cdot Z^{(n)}(z_0; z),$$

and similarly for the $\tilde{Z}^{(n)}$'s. Also, we shall write Σ-monogenic instead of (Σ, N)-monogenic, and use j in place of j_N.

At first glance, the iterated Σ- and Σ'-integration of the $2N$ constants $1, j, j^2, \cdots j^{2N-1}$ might be expected to yield $2N$ different functions as the result of each integration operation. However, each time, essentially only two new functions appear, because

$$
\begin{aligned}
&j^{2k} \cdot Z^{(n)}(z_0; z) = j^{2k} Z^{(n)}(z_0; z), \\
(7.4)\qquad &j^{2k+1} \cdot Z^{(n)}(z_0; z) = j^{2k}[j \cdot Z^{(n)}(z_0; z)],
\end{aligned}
$$

and

$$
\begin{aligned}
&j^{2k} \cdot \tilde{Z}^{(n)}(z_0; z) = j^{2k} \tilde{Z}^{(n)}(z_0; z), \\
(7.5)\qquad &j^{2k+1} \cdot \tilde{Z}^{(n)}(z_0; z) = j^{2k}[j \cdot \tilde{Z}^{(n)}(z_0; z)]
\end{aligned}
$$

where $k=0,1,2,\cdots$. Obviously, (7.4) and (7.5) hold for $n=0$. And if they hold for $n-1$, then they also hold for n, since in view of (7.1) and (6.6)

$$\begin{aligned} j^{2k}\cdot Z^{(n)}(z_0;z) &= n\int_{z_0}^{z}(j^{2k}\cdot \bar{Z}^{(n-1)}(z_0;z))\,d_{\Sigma'}z \\ &= n\int_{z_0}^{z} j^{2k}\bar{Z}^{(n-1)}(z_0;z)\,d_{\Sigma'}z \\ &= j^{2k}n\int_{z_0}^{z}\bar{Z}^{(n-1)}(z_0;z)\,d_{\Sigma'}z \\ &= j^{2k}Z^{(n)}(z_0;z), \end{aligned}$$

which verifies the first equation of (7.4). A similar reasoning yields the second equation of (7.4). The two equations of (7.5) may be derived analogously, or obtained at once by noticing that if the formal powers formed with respect to the matrix Σ' are denoted by $a\cdot Z'^{(n)}$, etc., we have

$$\begin{aligned} a\cdot Z'^{(n)}(z_0;z) &= a\cdot \bar{Z}^{(n)}(z_0;z), \\ a\cdot \bar{Z}'^{(n)}(z_0;z) &= a\cdot Z^{(n)}(z_0;z), \end{aligned} \tag{7.6}$$

and that (7.4) written with respect to Σ' is (7.5).

Clearly, for $a=\sum_{i=0}^{2N-1}\alpha_i j^i$, the α_i being real constants,

$$a\cdot Z^{(n)}(z_0;z)=\sum_{i=0}^{2N-1}\alpha_i[j^i\cdot Z^{(n)}(z_0;z)], \tag{7.7}$$

and similarly for $a\cdot\bar{Z}^{(n)}(z_0;z)$.

With the aid of (7.4), we obtain, from (7.7), the formula

$$\begin{aligned} [\sum_{i=0}^{2N-1}\alpha_i j^i]\cdot Z^{(n)}(z_0;z) &= [\sum_{k=0}^{N-1}\alpha_{2k}j^{2k}]Z^{(n)}(z_0;z) \\ &\quad + [\sum_{k=0}^{N-1}\alpha_{2k+1}j^{2k}][j\cdot Z^{(n)}(z_0;z)], \end{aligned} \tag{7.8}$$

which gives $a\cdot Z^{(n)}$ explicitly, once $Z^{(n)}$ and $j\cdot Z^{(n)}$ are known.

We call $a\cdot Z^{(n)}$ the "formal product" of "a" and the "formal power" $Z^{(n)}$.

B. If $f(z)=\sum_{i=0}^{2N-1}a_i j^i$, $a_i=a_i(x,y)$, is Σ-monogenic, then the a_i satisfy $E(\Sigma,N)$. Consequently, the formal powers $a\cdot Z^{(n)}(z_0;z)$ furnish us an infinite number of particular solutions of the system $E(\Sigma,N)$. In order to

further identify these particular solutions, and to obtain formulas which are formally equivalent to the binomial formula for $a(z-z_0)^n$, we use the following real functions (1): [See (1.10)]

$$X^{(0)}(x_0;x)=X^{*(0)}(x_0;x)=Y^{(0)}(y_0;y)=Y^{*(0)}(y_0;y)\equiv 1$$

$$(7.9)\qquad \begin{aligned}
X^{(n)}(x_0;x) &= \begin{cases} n\displaystyle\int_{x_0}^{x} X^{(n-1)}dX & \text{if } n \text{ is odd,}\\ n\displaystyle\int_{x_0}^{x} X^{(n-1)}dX^{*} & \text{if } n \text{ is even,}\end{cases}\\
X^{*(n)}(x_0;x) &= \begin{cases} n\displaystyle\int_{x_0}^{x} X^{*(n-1)}dX^{*} & \text{if } n \text{ is odd,}\\ n\displaystyle\int_{x_0}^{x} X^{*(n-1)}dX & \text{if } n \text{ is even,}\end{cases}\\
Y^{(n)}(y_0;y) &= \begin{cases} n\displaystyle\int_{y_0}^{y} Y^{(n-1)}dY & \text{if } n \text{ is odd,}\\ n\displaystyle\int_{y_0}^{y} Y^{(n-1)}dY^{*} & \text{if } n \text{ is even,}\end{cases}\\
Y^{*(n)}(y_0;y) &= \begin{cases} n\displaystyle\int_{y_0}^{y} Y^{*(n-1)}dY^{*} & \text{if } n \text{ is odd,}\\ n\displaystyle\int_{y_0}^{y} Y^{*(n-1)}dY & \text{if } n \text{ is even.}\end{cases}
\end{aligned}$$

In other words:

$$X^{(n)}(x_0;x)=n!\int_{x_0}^{x} 1/\sigma_1\int_{x_0}\sigma_2\cdots\int_{x_0}\sigma_2\int_{x_0} dx^n/\sigma_1 \qquad (n \text{ integrals, } n \text{ odd})$$

$$X^{(n)}(x_0;x)=n!\int_{x_0}^{x} \sigma_2\int_{x_0} 1/\sigma_1\cdots\int_{x_0}\sigma_2\int_{x_0} dx^n/\sigma_1 \qquad (n \text{ integrals, } n \text{ even})$$

and similarly for $X^{*(n)}$, $Y^{(n)}$, $Y^{*(n)}$. If $x_0=0$ (or $y_0=0$), we write simply $X^{(n)}(x)$, $X^{*(n)}(x)$, (or $Y^{(n)}(y)$, $Y^{*(n)}(y)$). It is to be noticed that the definition of these functions is independent of N.

Making use of the functions introduced in (7.9), we shall be able to obtain an analogue of the binomial formula for $(x+j_N y)^p$, by replacing the powers of x and y in (2.31) and (2.32) by the appropriate functions of (7.9). We have, for $q=0,1,2,\cdots$, in the notation of (2.33),

$$(7.10)\qquad Z^{(2q)}(z_0;z)=\sum_{i=0}^{2N-1}\Big[\sum_{l=0}^{2q} A_{i,l,2q}{}^{(N)}\,Y^{(l)}(y_0;y)X^{*(2q-l)}(x_0;x)\Big]j^i$$

$$(7.11)\qquad Z^{(2q+1)}(z_0;z)=\sum_{i=0}^{2N-1}\Big[\sum_{l=0}^{2q+1}A_{i,l,2q+1}{}^{(N)}Y^{(l)}(y_0;y)X^{(2q+1-l)}(x_0;x)\Big]j^i$$

$$(7.12)\qquad j\cdot Z^{(2q)}(z_0;z)=\sum_{i=0}^{2N-1}\Big[\sum_{l=0}^{2q}B_{i,l,2q}{}^{(N)}Y^{*(l)}(y_0;y)X^{(2q-l)}(x_0;x)\Big]j^i$$

$$(7.13)\qquad j\cdot Z^{(2q+1)}(z_0;z)=\sum_{i=0}^{2N-1}\Big[\sum_{l=0}^{2q+1}B_{i,l,2q+1}{}^{(N)}Y^{*(l)}(y_0;y)\,X^{*(2q+1-l)}(x_0;x)\Big]j^i.$$

Equation (7.10) may be verified by induction, employing the binomial identity (2.29). (7.11) is then obtained by Σ- and Σ'-differentiating $\dfrac{1}{(2q+2)(2q+1)}Z^{(2q+2)}(z_0;z)$. (7.12) and (7.13) follow immediately, after observing that if we introduce the formal powers $a\cdot Z,'^{(n)}$, etc., which are formed with respect to the matrix $\Sigma,'$, we have

$$j\cdot Z^{(n)}(z_0;z)=jZ,'^{(n)}(z_0;z).$$

C. It is possible to obtain a simple inequality for the absolute value of the formal powers. We have

$$(7.14)\qquad \|\alpha\cdot Z^{(n)}(z_0;z)\|\leq 2\,\|\alpha\|\,\Gamma_N[2^{\frac{1}{2}}\Gamma_N m_\Sigma(|z_0|+|z|)\,|z-z_0|]^n,$$

where $\alpha=\sum_{i=0}^{2N-1}\alpha_i j^i$ is a hypercomplex constant, Γ_N is the constant occurring in (2.26), and $|z_0|^2=x_0^2+y_0^2$, $|z|^2=x^2+y^2$.

First, we obtain inequalities for $Z^{(n)}(z_0;z)$ and for $j\cdot Z^{(n)}(z_0;z)$. Using the non-decreasing function $m_\Sigma(A)$, defined by

$$m_\Sigma(A)=\max_{i=1,2}\ \max_{l=\pm1}\ \max_{|t|\leq A}\{\sigma_i(t)^l,\tau_i(t)^l\},$$

we have that [see (1), page 75]

$$(7.15)\qquad m_\Sigma{}^{-n}(|x_0|+|x|)\leq\frac{X^{(n)}(x_0;x)}{(x-x_0)^n}\leq m_\Sigma{}^n(|x_0|+|x|),$$

with similar inequalities for X^*, Y and Y^*. We also know [see (2.31) and (2.32)] that

$$(7.16)\qquad Z^{(n)}(z_0;z)=\begin{cases}\displaystyle\sum_{i=0}^{n}\binom{n}{i}X^{(n-i)}Y^{(i)}j^i & \text{if } n \text{ is odd,}\\ \displaystyle\sum_{i=0}^{n}\binom{n}{i}X^{*(n-i)}Y^{(i)}j^i & \text{if } n \text{ is even,}\end{cases}$$

$$j\cdot Z^{(n)}(z_0;z)=\begin{cases}\displaystyle j\sum_{i=0}^{n}\binom{n}{i}X^{*(n-i)}Y^{*(i)}j^i & \text{if } n \text{ is odd,}\\ \displaystyle j\sum_{i=0}^{n}\binom{n}{i}X^{(n-i)}Y^{*(i)}j^i & \text{if } n \text{ is even.}\end{cases}$$

Consequently,

$$(7.17)\qquad \left.\begin{array}{l}\| Z^{(n)}(z_0;z)\| \\ \| j\cdot Z^{(n)}(z_0;z)\|\end{array}\right\} \leqq 2^{n/2}\Gamma_N{}^n m_\Sigma{}^n(|z_0|+|z|)|z-z_0|^n,$$

since by (2.26), (2.24), (2.25), plus (2.27) with $\alpha=j$, we have

$$\| j\cdot Z^{(n)}(z_0;z)\| \leqq \begin{cases}\Gamma_N \| j\| \ \| \sum_{i=0}^{n}\binom{n}{i} X^{*(n-i)}Y^{*(i)}j^i \| \\ \Gamma_N \| j\| \ \| \sum_{i=0}^{n}\binom{n}{i} X^{(n-i)}Y^{*(i)}j^i \|\end{cases}$$

$$\leqq \Gamma_N \sum_{i=0}^{n}\binom{n}{i} m_\Sigma{}^n(|z_0|+|z|)|x-x_0|^{n-i}|y-y_0|^i\Gamma_N{}^{i-1}$$

$$\leqq m_\Sigma{}^n(|z_0|+|z|)\Gamma_N{}^n[|x-x_0|+|y-y_0|]^n$$

$$\leqq 2^{n/2}\Gamma_N{}^n m_\Sigma{}^n(|z_0|+|z|)|z-z_0|^n.$$

The first inequality of (7.17) may be verified similarly.

But from (7.8) and (2.27)

$$\| \alpha\cdot Z^{(n)}(z_0;z)\| \leqq \Gamma_N \| \sum_{k=0}^{N-1}\alpha_{2k}j^{2k}\| \ \| Z^{(n)}(z_0;z)\|$$

$$(7.18)\qquad + \Gamma_N \| \sum_{k=0}^{N-1}\alpha_{2k+1}j^{2k}\| \ \| j\cdot Z^{(n)}(z_0;z)\|$$

$$\leqq \Gamma_N \|\alpha\| \,[\| Z^{(n)}(z_0;z)\| + \| j\cdot Z^{(n)}(z_0;z)\|],$$

Equation (7.14) then follows from (7.17) and (7.18).

8. Null functions. There are certain (Σ,N)-monogenic functions $f(z)=\sum_{i=0}^{2N-1} a_i j_N{}^i$ such that one or more of the real functions a_i vanish identically. This section will be devoted to the determination of all such "null functions."

THEOREM 8.1. *Let D be a domain containing the point z_0. If $f(z)=\sum_{i=0}^{q} a_i j_N{}^i$ where $0\leqq q\leqq 2N-2$, is (Σ,N)-monogenic in D, then*

$$(8.1)\qquad f(z)=\sum_{i=0}^{q}\alpha_i\cdot Z^{(i)}(z_0;z),$$

where

$$(8.2)\qquad \alpha_i=\sum_{l=0}^{q-i}\alpha_{il}j_N{}^l,$$

$(i=0,\cdots,q)$ are hypercomplex constants. (Conversely, any function defined by (8.1) and (8.2) is of the form $f(z)=\sum_{i=0}^{2N-1} a_i j_N{}^i$, where $a_{q+1}=a_{q+2}=\cdots a_{2N-1}\equiv 0$).

Proof. Given any matrix Σ, the conclusion is true for $q=0$. For then $f(z)$ is merely a real constant. Suppose that for any matrix Σ the assertion is true for q and let

$$f(z)=\sum_{i=0}^{2N-1} a_i j_N{}^i, \qquad a_{q+2}=a_{q+3}=\cdots=a_{2N-1}\equiv 0, \tag{8.3}$$

be (Σ, N)-monogenic in D. Then

$$\begin{aligned} a_{2k,X}&=a_{2k+1,Y},\\ a_{2k,Y^*}&=a_{2k-1,X^*}, \end{aligned} \tag{8.4}$$

$(k=0,\cdots,N-1)$. If $q+1=2k$ for some k, then, in view of the first equation of (8.4),

$$a_{q+1,X}=a_{q+2,Y}=0.$$

If $q+1=2k+1$ for some k, then, in view of the second equation of (8.4),

$$0=a_{q+2,Y^*}=a_{q+1,X^*}.$$

From Definition 3.2, we see that

$$\frac{d_{(\Sigma,N)}f}{d_{(\Sigma,N)}z}=\sum_{i=0}^{q} b_i j_N{}^i. \tag{8.5}$$

The function $\dfrac{d_{(\Sigma,N)}f}{d_{(\Sigma,N)}z}$ is (Σ', N)-monogenic; hence, by virtue of the induction hypothesis, (8.5) and (8.1) give

$$\frac{d_{(\Sigma,N)}f}{d_{(\Sigma,N)}z}=\sum_{i=0}^{q}\beta_i\cdot\bar{Z}^{(i)}(z_0;z), \tag{8.6}$$

where

$$\beta_i=\sum_{l=0}^{q-i}\beta_{il}j_N{}^l, \tag{8.7}$$

$(i=0,\cdots,q)$ are hypercomplex constants.

Equation (8.6) may be rewritten as follows:

$$\frac{d_{(\Sigma,N)}f}{d_{(\Sigma,N)}z}=\sum_{i=1}^{q+1}\beta_{i-1}\cdot\bar{Z}^{(i-1)}(z_0;z)=\sum_{i=1}^{q+1}i\alpha_i\cdot\bar{Z}^{(i-1)}(z_0;z), \tag{8.8}$$

by defining

$$\alpha_i=\frac{\beta_{i-1}}{i}, \tag{8.9}$$

$(i=1,\cdots,q+1)$. Setting

$$\alpha_{il}=\frac{\beta_{i-1,l}}{i},$$

for $(i=1,\cdots,q+1)$, $(l=0,\cdots,q-i)$, gives

$$\alpha_i=\sum_{l=0}^{q+1-i}\alpha_{il}j_N{}^l, \tag{8.10}$$

$(i=1,\cdots,q+1)$.

The function $f(z)$ of (8.3) is obtained by (Σ',N)-integrating (8.8). [This (Σ',N)-integration is possible whether D is simply connected or not, since the function given by (8.8) is (Σ',N)-monogenic at all points of the plane.] The result is

$$f(z)=\sum_{i=1}^{q+1}\alpha_i\cdot Z^{(i)}(z_0;z)+\alpha_0=\sum_{i=0}^{q+1}\alpha_i\cdot Z^{(i)}(z_0;z), \tag{8.11}$$

where

$$\alpha_0=\sum_{l=0}^{q+1}\alpha_{0l}j_N{}^l, \tag{8.12}$$

is a hypercomplex constant of a nature determined by the condition (8.3) on the functions a_i. The last three equations serve to verify the conclusion for $q+1$.

LEMMA 8.1. *For each $n\geqq 1$, let E_n be the set of $2n$ functions*

$$\begin{aligned} F_{2k}&=X^{(2k)},\\ F_{2k+1}&=X^{*(2k+1)}, \end{aligned} \tag{8.13}$$

$(k=0,\cdots,n-1)$. E_n is a linearly independent set of functions on any interval $a<x<b$.

The proof is by induction.

THEOREM 8.2a. *Let D be a domain and z_0 be a point of D. If $f(z)=\sum_{i=0}^{2N-1}a_ij_N{}^i$ is (Σ,N)-monogenic in D and $a_{2k}\equiv 0$, for some fixed k, where $0\leq k\leq N-1$, then*

$$f(z)=1/K_{N,k}\sum_{i=0}^{2N-2}\alpha_i\cdot Z^{(i)}(z_0;z), \tag{8.14}$$

where

$$\alpha_i=\sum_{l=1}^{2N-1-i}\alpha_{il}j_N{}^l, \tag{8.15}$$

$(i=0,1,\cdots,2N-2)$ are hypercomplex constants and the $K_{N,k}$ are given by Theorem 2.7. [Conversely, any function defined by (8.14) and (8.15) is of the form $f(z)=\sum_{i=0}^{2N-1}a_ij_N{}^i$, where $a_{2k}\equiv 0$.]

Proof. The function

$$(8.16)\qquad F(z)=\sum_{i=0}^{2N-1}A_ij_N{}^i=K_{N,k}f(z),$$

is (Σ, N)-monogenic and $A_0=a_{2k}=0$. In view of the second equation of $E(\Sigma, N)$, $0=A_{0,Y^*}=A_{2N-1,X^*}$. Hence, the (Σ', N)-monogenic function $\dfrac{d_{(\Sigma,N)}F}{d_{(\Sigma,N)}z}$ has the form

$$(8.17)\qquad \frac{d_{(\Sigma,N)}F}{d_{(\Sigma,N)}z}=\sum_{i=0}^{2N-2}A_i{}^{[1]}j_N{}^i.$$

By Theorem 8.1, we have

$$(8.18)\qquad \frac{d_{(\Sigma,N)}F}{d_{(\Sigma,N)}z}=\sum_{i=0}^{2N-2}\beta_i\cdot\tilde{Z}^{(i)}(z_0;z),$$

where

$$(8.19)\qquad \beta_i=\sum_{l=0}^{2N-2-i}\beta_{il}j_N{}^l,$$

$(i=0,\cdots,2N-2)$ are hypercomplex constants.

In addition, $A_0{}^{[1]}=A_{0,X}=0$, since $A_0=0$. Recalling (7.8), and by virtue of (2.31) and (2.32), for $0\leq q\leq N-1$

$$(8.20)\qquad \begin{aligned}&Re[\tilde{Z}^{(2q)}(z_0;z)]=X^{(2q)},\\ &Re[j_N\cdot\tilde{Z}^{(2q)}(z_0;z)]=0,\end{aligned}$$

and for $0\leq q\leq N-2$

$$(8.21)\qquad \begin{aligned}&Re[\tilde{Z}^{(2q+1)}(z_0;z)]=X^{*(2q+1)},\\ &Re[j_N\cdot\tilde{Z}^{(2q+1)}(z_0;z)]=0,\end{aligned}$$

we obtain from (8.17) and (8.18), by equating the real part of $\dfrac{d_{(\Sigma,N)}F}{d_{(\Sigma,N)}z}$ to zero

$$(8.22)\qquad A_0{}^{[1]}=\sum_{h=0}^{N-1}\beta_{2h,0}X^{(2h)}+\sum_{h=0}^{N-2}\beta_{2h+1,0}X^{*(2h+1)}=0.$$

By Lemma 8.1, (8.22) implies

$$\beta_{i,0}=0,$$

$(i=0,1,2,\cdots,2N-2)$. Consequently, $\beta_{2N-2}=0$ in (8.19) and equations (8.18) and (8.19) become

$$(8.23)\qquad \frac{d_{(\Sigma,N)}F}{d_{(\Sigma,N)}z}=\sum_{i=0}^{2N-3}\beta_i\cdot\tilde{Z}^{(i)}(z_0;z),$$

and

$$\beta_i = \sum_{l=1}^{2N-2-i} \beta_{il} j_N{}^l, \tag{8.24}$$

$(i = 0, \cdots, 2N - 3)$.

Let

$$\alpha_i = \frac{\beta_{i-1}}{i} = 1/i \sum_{l=1}^{2N-2-(i-1)} \beta_{i-1,l} j_N{}^l, \tag{8.25}$$

$(i = 1, \cdots, 2N - 2)$, and

$$\alpha_{il} = \beta_{i-1,l}/i, \tag{8.26}$$

for $i = 1, \cdots, 2N - 2$, $l = 1, \cdots, 2N - 1 - i$. Then, from (8.24), (8.25), and (8.26)

$$\alpha_i = \sum_{l=1}^{2N-1-i} \alpha_{il} j_N{}^l, \tag{8.27}$$

$(i = 1, \cdots, 2N - 2)$.

Using (8.25) and (8.27), equation (8.23) may be rewritten thus

$$\frac{d_{(\Sigma,N)} F}{d_{(\Sigma,N)} z} = \sum_{i=1}^{2N-2} \beta_{i-1} \cdot \bar{Z}^{(i-1)}(z_0; z) = \sum_{i=1}^{2N-2} i\alpha_i \cdot \bar{Z}^{(i-1)}(z_0; z). \tag{8.28}$$

$F(z)$ is found by (Σ', N)-integrating (8.28).

$$F(z) = \sum_{i=1}^{2N-2} \alpha_i \cdot \bar{Z}^{(i)}(z_0; z) + \alpha_0 = \sum_{i=0}^{2N-2} \alpha_i \cdot Z^{(i)}(z_0; z), \tag{8.29}$$

where

$$\alpha_0 = \sum_{l=1}^{2N-1} \alpha_{0l} j_N{}^l, \tag{8.30}$$

is a hypercomplex constant of a nature determined by the condition that $A_0 = 0$ in (8.16). Equations (8.29) and (8.16) yield (8.14), and equations (8.30) and (8.27) yield (8.15).

By a similar reasoning, we obtain

THEOREM 8.2b. *Let D be a domain and let z_0 be a point of D. If $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ is (Σ, N)-monogenic in D and $a_{2k+1} \equiv 0$ for some fixed k, where $0 \leq k \leq N - 1$, then*

$$f(z) = 1/K_{N,k}\Big[\sum_{i=1}^{2N-2} \alpha_i \cdot Z^{(i)}(z_0; z) + \alpha_0\Big], \tag{8.31}$$

where

$$\alpha_i = \sum_{l=2}^{2N-i} \alpha_{il} j_N{}^l, \tag{8.32}$$

$(i=1,\cdots,2N-2)$ *are hypercomplex constants,* $\alpha_0=\sum_{l=0}^{2N-1}\alpha_{0l}j_N{}^l$ *is a hypercomplex constant such that* $\alpha_{01}=0$, *and* $K_{N,\Sigma}$ *is given by Theorem* 2.7. *(Conversely, any function defined by* (8.31) *and* (8.32) *is of the form* $f(z)=\sum_{i=0}^{2N-1}a_ij_N{}^i$, *where* $a_{2k+1}\equiv 0$.)

THEOREM 8.3. *If* $f(z)=\sum_{i=0}^{2N-1}a_ij_N{}^i$ *is* (Σ,N)*-monogenic in a domain* D *and one of the functions* a_i *vanishes identically in* D, *then* $f(z)$ *is defined, single-valued, and* (Σ,N)*-monogenic over the whole plane.*

Proof. In D, $f(z)$ must coincide with either a function of the form (8.14) or a function of the form (8.31), and has the desired properties. Since the "components" a_i of a (Σ,N)-monogenic function are analytic functions of x and y, they are uniquely determined by their values in any neighborhood.

9. (Σ,N)-monogenic functions with prescribed a_i. If $f(z)=u+iv$ is an analytic function of a complex variable $x+iy$, then u and v, the real and imaginary parts of $f(z)$, satisfy Laplace's equation. The analogue of this proposition for (Σ,N)-monogenic functions was considered in **5**. Again, given a harmonic function $u=u(x,y)$ there exist two analytic functions of $x+iy$, one having u as its real part and the other having u as its imaginary part. This section will be concerned with the analogue of this "inverse" problem for (Σ,N)- monogenic functions. (A special case has already been treated in **8**.)

To facilitate later computations, we note that

$$(9.1)\qquad \begin{aligned} M^Nu&=\left(\frac{\partial^2}{\partial X\partial X^*}+\frac{\partial^2}{\partial Y^*\partial Y}\right)^n u=\sum_{k=0}^{n}\binom{n}{k}\frac{\partial^{2n}u}{(\partial X\partial X^*)^k(\partial Y^*\partial Y)^{n-k}},\\ L^Nu&=\left(\frac{\partial^2}{\partial X^*\partial X}+\frac{\partial^2}{\partial Y\partial Y^*}\right)^n u=\sum_{k=0}^{n}\binom{n}{k}\frac{\partial^{2n}u}{(\partial X^*\partial X)^k(\partial Y\partial Y^*)^{n-k}}.\end{aligned}$$

We treat first the case when the real part a_0 is prescribed.

LEMMA 9.1. *Suppose that* $u=(x,y)$ *of class* $C^{(2N)}$ *satisfies the equation* $L^Nu=0$ *in a domain* D. *Then, there exists a function* $f(z)=\sum_{i=0}^{2N-1}a_ij_N{}^i$ *which is* (Σ,N)*-monogenic in* D *and such that* $a_0=u$.

Proof. We shall prove our result first when D is a rectangle with sides parallel to the axes. For convenience, we suppose that the origin is an interior

point of D. The proof is by induction. The conclusion is true for $N = 1$ for any matrix Σ. In what follows, we suppose that the conclusion is true for $N - 1$ for any matrix Σ.

It follows that given a function $v = v(x, y)$ of class $C^{(2N-2)}$ satisfying $M^{N-1}v = 0$ in D, there exists a function $g(z) = \sum_{i=0}^{2N-3} b_i j_{N-1}{}^i$ which is $(\Sigma, N-1)$-monogenic in D and such that $b_1 = v$. Because for the matrix $\Sigma_,'$ of (6.1), we have

$$L_,' = v_{X^*X} + v_{YY^*} = Mv$$

and consequently $L_,'^{N-1}v = M^{N-1}v = 0$ for the given function v. Hence by the induction hypothesis, there exists a function $f(z) = \sum_{i=0}^{2N-3} a_i j_{N-1}{}^i$ which is $(\Sigma_,', N-1)$-monogenic in D and such that $a_0 = v$. From Theorem 6.1, it follows that $\sum_{i=0}^{2N-1} b_i j_{N-1}{}^i = g(z) = j_{N-1}f(z)$ is $(\Sigma, N-1)$-monogenic, and from Theorem 2.6 that $b_1 = a_0 = v$.

Now, let $u = u(x, y)$ of class $C^{(2N)}$ satisy $L^N u = 0$ in D. Writing $a_0 = u$, we seek to construct $2N - 1$ functions $a_0, \cdots, a_{2N-1}$ such that $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ is (Σ, N)-monogenic in D. In order to do this, we shall introduce three auxiliary functions $\omega_1(x, y)$, $\omega_2(x, y)$, and $\phi(x, y)$; define $a_1(x, y)$ and $a_{2N-1}(x, y)$; and from the three functions a_0, a_1, and a_{2N-1} deduce the rest. The restriction that D be a rectangle appears in the definition of a_1 and a_{2N-1}.

Recalling that $\int \cdots dX^* = \int \cdots \sigma_2(x)dx$ and $\int \cdots dY = \int \cdots (1/\tau_1(y))dy$, we set

$$\omega_1(x, y) \equiv \int_0^y \frac{\partial a_0}{\partial X}\, dY, \tag{9.2}$$

$$\omega_2(x, y) \equiv \int_0^x \frac{\partial a_0}{\partial Y^*}\, dX^*,$$

and

$$\phi(x, y) \equiv M^{N-1}(\omega_1 + \omega_2). \tag{9.3}$$

It will be shown immediately that

$$\phi(x, y) = \lambda(x) + \psi(y). \tag{9.4}$$

Since (9.2) yields

$$\frac{\partial \omega_1}{\partial Y} = \frac{\partial a_0}{\partial X}, \qquad \frac{\partial \omega_2}{\partial X^*} = \frac{\partial a_0}{\partial Y^*},$$

we may use the identities

$$\frac{\partial^{2p}}{(\partial Y^*\partial Y)^p} = \frac{\partial}{\partial Y^*}\,\frac{\partial^{2p-2}}{(\partial Y\partial Y^*)^{p-1}}\,\frac{\partial}{\partial Y},$$

$$\frac{\partial^{2p}}{(\partial X\partial X^*)^p}\,\frac{\partial}{\partial X} = \frac{\partial}{\partial X}\,\frac{\partial^{2p}}{(\partial X^*\partial X)^p},$$

to obtain from (9.2)

$$\begin{aligned} M^{N-1}\omega_1 &= \sum_{k=0}^{N-1}\binom{N-1}{k}\frac{\partial^{2N-2}\omega_1}{(\partial X\partial X^*)^k(\partial Y^*\partial Y)^{N-1-k}} \\ &= \frac{\partial^2}{\partial X\partial Y^*}\sum_{k=0}^{N-2}\binom{N-1}{k}\frac{\partial^{2N-4}a_0}{(\partial X^*\partial X)^k(\partial Y\partial Y^*)^{N-2-k}} \\ &\quad+\int_0^y \frac{\partial^{2N-1}a_0}{\partial X(\partial X^*\partial X)^{N-1}}\,dY. \end{aligned}$$

Performing a similar simplification on $M^{N-1}\omega_2$, we obtain from (9.3)

$$\begin{aligned} \phi(x,y) &= \frac{\partial^2}{\partial X\partial Y^*}\sum_{k=0}^{N-2}\binom{N-1}{k}\frac{\partial^{2N-4}a_0}{(\partial X^*\partial X)^k(\partial Y\partial Y^*)^{N-2-k}} \\ &\quad+\int_0^y \frac{\partial^{2N-1}a_0}{\partial X(\partial X^*\partial X)^{N-1}}\,dY \\ &\quad+\frac{\partial^2}{\partial X\partial Y^*}\sum_{k=1}^{N-1}\binom{N-1}{k}\frac{\partial^{2N-4}a_0}{(\partial X^*\partial X)^{k-1}(\partial Y\partial Y^*)^{N-1-k}} \\ &\quad+\int_0^x \frac{\partial^{2N-1}a_0}{\partial Y^*(\partial Y\partial Y^*)^{N-1}}\,dX^*. \end{aligned}$$

Hence

$$\begin{aligned} \phi_{X^*Y} &= \sum_{k=0}^{N-2}\binom{N-1}{k}\frac{\partial^{2N}a_0}{(\partial X^*\partial X)^{k+1}(\partial Y\partial Y^*)^{N-1-k}} \\ &\quad+\frac{\partial^{2N}a_0}{(\partial X^*\partial X)^N} \\ &\quad+\sum_{k=1}^{N-1}\binom{N-1}{k}\frac{\partial^{2N}a_0}{(\partial X^*\partial X)^k(\partial Y\partial Y^*)^{N-k}} \\ &\quad+\frac{\partial^{2N}a_0}{(\partial Y\partial Y^*)^N} \\ &= \sum_{k=1}^{N-1}\binom{N-1}{k-1}\frac{\partial^{2N}a_0}{(\partial X^*\partial X)^k(\partial Y\partial Y^*)^{N-k}} \\ &\quad+\frac{\partial^{2N}a_0}{(\partial X^*\partial X)^N} \\ &\quad+\sum_{k=1}^{N-1}\binom{N-1}{k}\frac{\partial^{2N}a_0}{(\partial X^*\partial X)^k(\partial Y\partial Y^*)^{N-k}} \end{aligned}$$

$$
\begin{aligned}
&+\frac{\partial^{2N}a_0}{(\partial Y\partial Y^*)^N}\\
&=\sum_{k=0}^{N}\binom{N}{k}\frac{\partial^{2N}a_0}{(\partial X^*\partial X)^k(\partial Y\partial Y^*)^{N-k}}\\
&=L^Na_0=0.
\end{aligned}
$$

by (9.2) and the hypothesis concerning a_0. Since $\phi_{X^*Y}=(\tau_1/\sigma_2)\phi_{xy}$ and $\phi_{X^*Y}=0$, we have $\phi_{xy}=0$ and (9.3) follows.

Now, let the functions $A(x)$ and $B(y)$ be defined by the equations

$$\frac{d^{2N-2}A(x)}{(dXdX^*)^{N-1}}=-\lambda(x), \tag{9.5}$$

and

$$\frac{d^{2N-2}B(y)}{(dY^*dY)^{N-1}}=-\psi(y). \tag{9.6}$$

It is easily seen that such real functions exist, although they are not uniquely determined.

It is now possible to define $a_1(x,y)$ and $a_{2N-1}(x,y)$. The system $E(\Sigma,N)$

$$
\begin{aligned}
&a_{2k,X}=a_{2k+1,Y},\\
&a_{2k,Y^*}=a_{2k-1,X^*}-a_{2N-1,X^*}\binom{N}{k},
\end{aligned}
\tag{9.7}
$$

$(k=0,\cdots,N-1)$ gives for $k=0$

$$
\begin{aligned}
&a_{0,X}=a_{1,Y},\\
&a_{0,Y^*}=-a_{2N-1,X^*},
\end{aligned}
\tag{9.8}
$$

In order to satisfy (9.8), we define [see (9.2)]

$$
\begin{aligned}
&a_1(x,y)\equiv\omega_1(x,y)+A(x),\\
&a_{2N-1}(x,y)\equiv-\omega_2(x,y)-B(y).
\end{aligned}
\tag{9.9}
$$

We have that

$$M^{N-1}(a_1-a_{2N-1})=0, \tag{9.10}$$

for, using (9.9)

$$
\begin{aligned}
M^{N-1}(a_1-a_{2N-1})&=M^{N-1}\omega_1+\frac{d^{2N-2}A(x)}{(dXdX^*)^{N-1}}\\
&\quad+M^{N-1}\omega_2+\frac{d^{2N-2}B(y)}{(dY^*dY)^{N-1}}\\
&=\phi-\lambda(x)-\psi(y)\\
&=0,
\end{aligned}
$$

by (9.3) and (9.4).

Since, in view of (9.10), we have one function satisfying $M^{N-1}v = 0$, we may use equation (4.1) "backwards" to produce the missing functions $a_2, \cdots, a_{2N-2}$. By the remark made at the beginning of the proof, there exists a $(\Sigma, N-1)$-monogenic function $g(z) = \sum_{i=0}^{2N-3} \alpha_i j_{N-1}{}^i$ such that $\alpha_1 = a_1 - a_{2N-1}$. Define $a_{2N-2}(x, y)$ by the relation

$$a_{2N-2} = a_0 - \alpha_0.$$

Since a_{2N-2} and a_{2N-1} are known, we may define a_{2k} and a_{2k+1} for $k = 1, \cdots, N-2$ by the relations [see equation (4.1)]

$$\begin{aligned} a_{2k} &= \alpha_{2k} + \binom{N-1}{k} a_{2N-2}, \\ a_{2k+1} &= \alpha_{2k+1} + \binom{N-1}{k} a_{2N-1}. \end{aligned} \tag{9.11}$$

Equation (9.11) also holds for $k = 0$, and for $k = N-1$ provided α_{2N-2} and α_{2N-1} are taken as zero. We must now show that the a_i thus defined satisfy (9.7). In order to do this, we remark that the α_i satisfy $E(\Sigma, N-1)$

$$\begin{aligned} \alpha_{2k,X} &= \alpha_{2k+1,Y}, \\ \alpha_{2k,Y^*} &= \alpha_{2k-1,X^*} - \alpha_{2N-3,X^*} \binom{N-1}{k}, \end{aligned} \tag{9.12}$$

$(k = 0, \cdots, N-2)$.

First, as has already been pointed out, in view of (9.9), the pair of equations obtained by setting $k = 0$ in (9.7) are satisfied. That is

$$\begin{aligned} a_{0,X} &= a_{1,Y}, \\ a_{0,Y^*} &= -a_{2N-1,X^*}. \end{aligned} \tag{9.13}$$

On the other hand, (9.12) gives for $k = 0$

$$\begin{aligned} \alpha_{0,X} &= \alpha_{1,Y}, \\ \alpha_{0,Y^*} &= -\alpha_{2N-3,X^*}, \end{aligned} \tag{9.14}$$

and using (9.11) for $k = 0$ and $k = N-2$, (9.14) becomes

$$\begin{aligned} a_{0,X} - a_{2N-2,X} &= a_{1,Y} - a_{2N-1,Y}, \\ a_{0,Y^*} - a_{2N-2,Y^*} &= -[a_{2N-3,X^*} - a_{2N-1,X^*} \binom{N-1}{N-2}], \end{aligned} \tag{9.15}$$

which, in view of (9.13), simplifies to

$$(9.16)\qquad \begin{aligned} &a_{2N-2,X} = a_{2N-1,Y},\\ &a_{2N-2,Y^*} = a_{2N-3,X^*} - a_{2N-1,X^*}\binom{N}{N-1}, \end{aligned}$$

the pair of equations obtained from (9.7) by setting $k = N-1$.

To prove that (9.9) holds for $k = 1, \cdots, N-2$, we first substitute into (9.12) the α_i's as given by (9.11). The result is

$$(9.17)\qquad \begin{aligned} &a_{2k,X} - \binom{N-1}{k} a_{2N-2,X} = a_{2k+1,Y} - \binom{N-1}{k} a_{2N-1,Y},\\ &a_{2k,Y^*} - \binom{N-1}{k} a_{2N-2,Y^*} = a_{2k-1,X^*} - \binom{N-1}{k-1} a_{2N-1,X^*}\\ &\qquad - \binom{N-1}{k}\left[a_{2N-3,X^*} - \binom{N-1}{N-2} a_{2N-1,X^*}\right], \end{aligned}$$

$(k = 1, \cdots, N-2)$. In view of (9.16), (9.17) becomes (9.7). Thus $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ is (Σ, N)-monogenic and $a_0 = u$ inside the rectangle D.

For convenience, if a domain D is such that Lemma 9.1 holds for D, then we shall call D an adequate domain. To obtain the lemma for a general *domain* D, we verify the following statements:

(1) If D_1 and D_2 are adequate domains and their intersection $D_1 \cdot D_2$ is a non-vacuous connected set, then their sum $D_1 + D_2$ is also an adequate domain.

(2) If D_1, D_2, and D_3 are adequate domains, $D_1 \cdot D_2$ and $D_2 \cdot D_3$ are non-vacuous connected sets, and D_1 and D_3 are mutually exclusive, then $D_1 + D_2 + D_3$ is an adequate domain.

(3) Let a domain be said to be of class n, $(n = 1, 2, \cdots)$, if its closure is the sum of n mutually exclusive rectangular interiors, plus their boundaries, the rectangular boundaries having their sides parallel to the x and y axes and such that if two of the boundaries have a common part, then the common part contains more than one point. For each n, every domain of class n is adequate.

(4) If a domain D is the sum of a monotone ascending sequence $D_1 \subset D_2 \subset D_3 \subset \cdots$ of adequate domains, then D is an adequate domain.

To prove (1), let $f_i(z)$, for $i = 1, 2$, be (Σ, N)-monogenic in D_i and such that $Re\, f_i(z) = u$ in D_i. $f_1(z)$ differs from $f_2(z)$ in $D_1 \cdot D_2$ by at most a null function (see Section 8), since $Re[f_1(z) - f_2(z)] = 0$ in $D_1 \cdot D_2$.

Hence, the function $g(z) \equiv f_1(z) - f_2(z)$ is defined over the whole plane (see Section 8), and the function

$$F(z) = \begin{cases} f_1(z) \text{ in } D_1, \\ f_2(z) + g(z) \text{ in } D_2, \end{cases}$$

is (Σ, N)-monogenic in $D_1 + D_2$ and $Re[F(z)] = u$ there. (2) follows by applying (1) first to D_1 and D_2 and then to $D_1 + D_2$ and D_3. (3) is proved by induction. A domain of class 1 is the interior of a rectangle; hence, it is adequate. It is easily verified that the closure of a domain of class n is the sum of the closure of a domain of class $n-1$ and the closure of a domain of class 1. The induction is completed by introducing a suitable rectangle [to play the role of D_2 in (2)] and reasoning as in the proof of (2). (4) follows by applying repeatedly the same argument used in the proof of (1) in order to extend the definition of a (Σ, N)-monogenic function $f_1(z)$ such that $Re\, f_1(z) = u$ in D_1 to all domains of the ascending sequence.

That Lemma 9.1 holds for any domain D follows from (4), since every domain is known to be the sum of a monotone ascending sequence of domains each of which belongs to class n for some n.

We remark that if D is simply connected, then the (Σ, N)-monogenic function obtained above is single-valued, by the analogue of the monodromy theorem for (Σ, N)-monogenic functions. If D is not simply connected, then the (Σ, N)-monogenic function obtained above may be multiple valued.

From Lemma 9.1, in the same way as was pointed out for $N-1$ in the course of the proof of that lemma, we have

LEMMA 9.2. *If $v = v(x, y)$ of class $C^{(2N)}$ satisfies the equation $M^N v = 0$ in a domain D, then there exists a (Σ, N)-monogenic function $g(z) = \sum_{i=0}^{2N-1} b_i j_N{}^i$ such that $b_1 = v$.*

The preceding considerations enable us to give a complete solution to the "inverse problem" for (Σ, N)-monogenic functions by means of the

THEOREM 9.1. *If $u = u(x, y)$ of class $C^{(2N)}$ is such that either (1) $L^N u = 0$ or (2) $M^N u = 0$ in a domain D, then for each integer $k = 0, \cdots, N-1$ there exists (1) a function $f_k(z) = \sum_{i=0}^{2N-1} a_{ki} j_N{}^i$ which is (Σ, N)-monogenic in D and such that $a_{k,2k} = u$ or (2) a function $g_k(z) = \sum_{i=0}^{2N-1} b_{ki} j_N{}^i$ which is (Σ, N)-monogenic in D and such that $b_{k,2k+1} = u$.*

We prove only (1), the proof for (2) being analogous. Let u satisfying $L^N u = 0$ be given and let $f_0(z) = \sum_{i=0}^{2N-1} a_{0i} j_N{}^i$ be the (Σ, N)-monogenic function given by Lemma 9.1, with $a_{00} = u$. Then $f_k(z) = \sum_{i=0}^{2N-1} a_{ki} j_N{}^i = (1/K_{N,k}) f_0(z)$, $k = 0, \cdots, N-1$ are the desired functions, where the $K_{N,k}$ are given in Theorem 2.7. By Theorem 2.7, $a_{k,2k} = a_{00} = u$. By Theorem 2.8 and Corollary 6.1, it follows that the $f_k(z)$ are (Σ, N)-monogenic.

Finally, we remark that to obtain the most general (Σ, N)-monogenic function $F_k(z) = \sum_{i=0}^{2N-1} A_{ki} j_N{}^i$ such that $A_{k,2k} = u$, we merely add to the $f_k(z)$ the appropriate "null functions" from **8**.

10. Formal power series and the expansion theorem. An expression of the form

$$\sum_{n=0}^{\infty} \alpha_n \cdot Z^{(n)}(z_0; z), \tag{10.1}$$

where $\alpha_n = \sum_{i=0}^{2N-1} \alpha_{ni} j_N{}^i$ are hypercomplex constants, will be called a formal power series. Our primary objective in this section is to show that if $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ is (Σ, N)-monogenic at a point z_0, then it can be represented uniquely by a formal power series (10.1) in a sufficiently small neighborhood of z_0. Preliminary to this, we shall establish the existence, under certain conditions on the coefficients α_n, of a domain of convergence of (10.1) and also show that a (Σ, N)-monogenic function in a domain D is uniquely determined by the value of the function and of all its (Σ, N)- and (Σ', N)-derivatives at a point of D.

In **3** the higher (Σ, N)-derivatives $f^{[n]}(z) = \sum_{i=0}^{2N-1} a_i{}^{[n]} j_N{}^i$ of a (Σ, N)-monogenic function $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ were defined, and it was shown that $f^{[2n]}(z)$ is (Σ, N)-monogenic and $f^{[2n+1]}(z)$ is (Σ', N)-monogenic.

Equation (3.5) gives that $f^{[1]}(z) = \dfrac{d_\Sigma f}{d_\Sigma z} = \sum_{i=0}^{2N-1} a_i{}^{[1]} j_N{}^i$, where

$$\begin{aligned} a_{2k}{}^{[1]} &= a_{2k,X} = \sigma_1 a_{2k,x}, \\ a_{2k+1}{}^{[1]} &= a_{2k+1,X^*} = (1/\sigma_2) a_{2k+1,x}, \end{aligned} \tag{10.2}$$

$(k=0,\cdots,N-1)$. Repeated application of (10.2) gives immediately

$$a_{2k}{}^{[2n]}=\frac{\partial^{2n}a_{2k}}{(\partial X^*\partial X)^n},\qquad a_{2k+1}{}^{[2n]}=\frac{\partial^{2n}a_{2k+1}}{(\partial X\partial X^*)^n},\tag{10.3}$$
$$a_{2k}{}^{[2n+1]}=\frac{\partial^{2n+1}a_{2k}}{\partial X(\partial X^*\partial X)^n},\qquad a_{2k+1}{}^{[2n+1]}=\frac{\partial^{2n+1}a_{2k+1}}{\partial X^*(\partial X\partial X^*)^n},$$

$(n=0,1,\cdots)$, $(k=0,\cdots N-1)$.

We shall employ a lemma given by Bers and Gelbart (1)

LEMMA 10.1. *Let $F(x)$, $G(x)$, $H(x)$ be analytic functions of the real variable x in the neighborhood of $x=0$. Define*

$$F_0(x)=F(x),\tag{10.4}$$

$$F_n(x)=\begin{cases}G(x)\ \dfrac{dF_{n-1}}{dx} & \text{if } n \text{ is odd,}\\[2ex] H(x)\ \dfrac{dF_{n-1}}{dx} & \text{if } n \text{ is even.}\end{cases}$$

Then there exists a constant $C>0$ such that

$$|F_n(0)|<n!\,C^n,\tag{10.5}$$

for all n.

The following lemma will be used in the proof of the expansion theorem.

LEMMA 10.2. *If $f(z)=\sum_{i=0}^{2N-1}a_ij_N{}^i$ is (Σ,N)-monogenic at $z=z_0$, then there exists a constant $C>0$ such that*

$$\|f^{[n]}(z_0)\|<n!\,C^n,\tag{10.6}$$

for all n.

Proof. Without loss in generality z_0 may be taken to be zero. Applying Lemma 10.1 for each $k=0,\cdots,N-1$, with

$$F(x)=a_{2k}(x,0),\qquad G(x)=\sigma_1(x),\qquad H(x)=1/\sigma_2(x),$$

and using (10.3), we see that there exist N constants C_{2k} such that

$$|a_{2k}{}^{[n]}(0,0)|<n!\,C_{2k}{}^n,\tag{10.7}$$

$(k=0,\cdots,N-1)$.

Similarly, there exist N constants C_{2k+1} such that

$$|a_{2k+1}{}^{[n]}(0,0)|<n!\,C_{2k+1}{}^n.\tag{10.8}$$

But, by (2.24) and (2.25)

$$\| f^{[n]}(0) \| \leqq \sum_{i=0}^{2N-1} | a_i^{[n]}(0,0) | \, \| j_N^i \| , \tag{10.9}$$

and $\| j_N^i \| = 1$ for $0 \leq i \leq 2N-1$, so that from (10.7) and (10.8)

$$\| f^{[n]}(0) \| < n! \, (\sum_{i=0}^{2N-1} C_i)^n. \tag{10.10}$$

THEOREM 10.1. *If* $f(z) = \sum_{i=0}^{2N-1} a_i j_N^i$ *is* (Σ, N)*-monogenic at* $z = z_0$ *and*

$$f^{[n]}(z_0) = 0, \qquad n = 0, 1, \cdots, \tag{10.11}$$

then

$$f(z) \equiv 0.$$

The proof is by induction on N. For $N = 1$ the theorem has been proved by Bers and Gelbart (1). Suppose that the conclusion is true for $N-1$, and let $f(z) = \sum_{i=0}^{2N-1} a_i j_N^i$ be a (Σ, N)-monogenic function satisfying (10.11) at $z = z_0$. Equation (10.11) implies that

$$a_i^{[n]}(z_0) = 0, \qquad n = 0, 1, \cdots, \tag{10.12}$$

$(i = 0, \cdots, 2N-1)$.

By Lemma 4.1, the function $\phi(z) = \sum_{i=0}^{2N-3} \alpha_i j_{N-1}^i$, where

$$\begin{aligned} \alpha_{2k} &= a_{2k} - \binom{N-1}{k} a_{2N-2}, \\ \alpha_{2k+1} &= a_{2k+1} - \binom{N-1}{k} a_{2N-1}, \end{aligned} \tag{10.13}$$

$(k = 0, \cdots, N-2)$, is $(\Sigma, N-1)$-monogenic in a neighborhood of z_0. From (10.11), we have that

$$\phi^{[n]}(z_0) = 0, \qquad n = 0, 1, \cdots, \tag{10.14}$$

since (10.13) and (10.3) together give

$$\begin{aligned} \alpha_{2k}^{[n]}(z_0) &= a_{2k}^{[n]}(z_0) - \binom{N-1}{k} a_{2N-2}^{[n]}(z_0) = 0, \\ \alpha_{2k+1}^{[n]}(z_0) &= a_{2k-1}^{[n]}(z_0) - \binom{N-1}{k} a_{2N-1}^{[n]}(z_0) = 0, \end{aligned} \tag{10.15}$$

$(k = 0, \cdots, N-2)$, using (10.12).

By (10.14) and the induction hypothesis, we have that $\phi(z)=\sum_{i=0}^{2N-3}\alpha_i j_{N-1}{}^i \equiv 0$, and from (10.13)

$$a_{2k}=\binom{N-1}{k}a_{2N-2},\qquad a_{2k+1}=\binom{N-1}{k}a_{2N-1},\tag{10.16}$$

$(k=0,\cdots,N-2)$. Since $f(z)=\sum_{i=0}^{2N-1}a_i j_N{}^i$ is (Σ,N)-monogenic, then

$$a_{2k,X}=a_{2k+1,Y},\qquad a_{2k,Y^*}=a_{2k-1,X^*}-a_{2N-1,X^*}\binom{N}{k},\tag{10.17}$$

$(k=0,\cdots,N-1)$. Replacing a_{2k} and a_{2k-1} in (10.17) by their values from (10.16) yields

$$a_{2N-2,X}=a_{2N-1,Y},\qquad a_{2N-2,Y^*}=-a_{2N-1,X^*}.\tag{10.18}$$

Replacing a_{2N-2} and a_{2N-1} in (10.18) by their values from (10.16) gives

$$a_{2k,X}=a_{2k+1,Y},\qquad a_{2k,Y^*}=-a_{2k+1,X^*},\tag{10.19}$$

$(k=0,\cdots,N-1)$. This last equation shows that the N functions

$$f_k(z)=a_{2k}+j_1a_{2k+1},\tag{10.20}$$

$(k=0,\cdots,N-1)$, are $(\Sigma,1)$-monogenic. But from (10.12) and (10.3), we have that

$$f_k^{[n]}(z_0)=0,\qquad n=0,1,\cdots\tag{10.21}$$

$(k=0,\cdots,N-1)$. Hence, $a_i=0$ for $i=0,\cdots,2N-1$ and $f(z)\equiv 0$

Returning to the consideration of the formal power series (10.1), we remark that such a series represents a (Σ,N)-monogenic function in any domain in which it converges uniformly, and may be (Σ,N)-integrated term by term in any such domain. We shall show that a domain of uniform convergence of (10.1) is determined by the sequence of coefficients α_n.

THEOREM 10.2. *Let $\alpha_n=\sum_{i=0}^{2N-1}\alpha_{ni}j_N{}^i$ be a sequence of numbers of A_{2N} and $z=z_0$ be fixed. If*

$$\frac{1}{\overline{\lim}\sqrt[n]{\|\alpha_n\|}}=R,\tag{10.22}$$

does not vanish, then there exists a neighborhood D of z_0 such that the formal power series

$$f(z) = \sum_{n=0}^{\infty} \alpha_n \cdot Z^{(n)}(z_0; z), \tag{10.23}$$

and all the formal power series obtained by successively (Σ, N)- and (Σ', N)-differentiating (10.23) term by term converge uniformly and absolutely in D.

Proof. Let D be the domain

$$|z - z_0| < \frac{\rho}{2^{\frac{1}{2}} \Gamma_N m_{\Sigma}(|z_0| + R)}, \qquad 0 < \rho < R. \tag{10.24}$$

From (10.22), it follows that D has the required property, for the inequality (7.14) implies

$$\sum_{n=0}^{M} \| \alpha_n \cdot Z^{(n)}(z_0; z) \| < 2\, \Gamma_N \sum_{n=0}^{M} \| \alpha_n \| \rho^n,$$

and similar inequalities hold for the derived series.

Successive termwise (Σ, N)- and (Σ', N)-differentiation of (10.23) and evaluation at z_0 of the resulting formal power series gives the "Taylor formulas"

$$f^{[n]}(z_0) = n!\, \alpha_n, \tag{10.25}$$

$n = 0, 1, \cdots$.

THEOREM 10.3. *Let $f(z) = \sum_{i=0}^{2N-1} a_i j_N{}^i$ be (Σ, N)-monogenic at $z = z_0$. Then $f(z)$ can be (uniquely) expanded at $z = z_0$ in a formal power series of the form* (10.23).

Proof. Let the sequence of hypercomplex numbers α_n be defined by (10.25). By virtue of Lemma 10.2

$$\overline{\lim}\ \sqrt[n]{\| \alpha_n \|} \leq C.$$

Hence, by Theorem 10.2, the formal power series

$$g(z) = \sum_{n=0}^{\infty} \alpha_n \cdot Z^{(n)}(z_0; z),$$

converges uniformly and absolutely in a neighborhood of z_0 and

$$g^{[n]}(z_0) = n!\, \alpha_n,$$

$n = 0, 1, \cdots$. Then $f(z) - g(z) \equiv 0$ by Theorem 10.1.

Theorems 10.1 and 10.3 serve to show that the class of functions $f(z) = \sum_{i=0}^{2N-1} a_i j_N^i$ which can be expanded in a formal power series of the form (10.1) in a neighborhood of z_0 coincides with the class of functions which are (Σ, N)-monogenic at z_0 according to Definition 3.1. This corresponds to the equivalence of the Cauchy and Weierstrass definitions of an analytic function.

11. Complete systems of particular solutions of $L^N u = 0$. We define

$$U_{Nn\nu}(x_0, y_0; x, y) = Re[j_N^\nu \cdot Z^{(n)}(z_0; z)], \tag{11.1}$$

where $z = x + iy$, $z_0 = x_0 + iy_0$, $n = 0, 1, \cdots, \nu = 0, 1, \cdots, 2N - 1$. Obviously, each U is a (real-valued) particular solution of the equation $L^N u = 0$ analytic in the whole plane. Notice that for $n < N$ some of the functions (11.1) vanish. The sequence $\{U_{Nn\nu}\}$ has the completeness properties mentioned in Section 1. We have

THEOREM 11.1. *Each solution $u = u(x, y)$ of the equation $L^N u = 0$ which is of class $C^{(2N)}$ in a domain containing the point z_0, admits the expansion*

$$u(x, y) = \sum_{n=0}^{\infty} \sum_{\nu=0}^{2N-1} a_{n\nu} U_{Nn\nu}(x_0, y_0; x, y), \tag{11.2}$$

where the a's are real constants and the series converges uniformly and absolutely in some neighborhood of z_0.

Proof. By Lemma 9.1, there exists, in a suitable neighborhood of z_0, a (Σ, N)-monogenic function $f(z) = \sum_{i=0}^{2N-1} a_i j_N^i$ such that $a_0 = u$. By Theorem 10.3, we have

$$f(z) = \sum_{n=0}^{\infty} \alpha_n \cdot Z^{(n)}(z_0; z), \tag{11.3}$$

where

$$\alpha_n = \sum_{\nu=0}^{2N-1} a_{n\nu} j_N^\nu,$$

$n = 0, 1, \cdots$, in a sufficiently small neighborhood of z_0. (11.2) follows upon equating the real parts of both sides of (11.3), using (7.8) and (11.1).

The sequence of real functions (11.1) depends upon the point z_0. In order to prove a more general expansion theorem, we establish

THEOREM 11.2. *Let z_0 and b be points of the plane and let α be a number of A_{2N}, then*

$$(11.4)\qquad \alpha \cdot Z^{(n)}(z_0; z) = \sum_{k=0}^{n} \binom{n}{k} A_k \cdot Z^{(n-k)}(b; z),$$

where

$$(11.5)\qquad A_k = \begin{cases} \alpha \cdot Z^{(k)}(z_0; b) & \text{if } k \text{ is even,} \\ \alpha \cdot \bar{\bar{Z}}^{(k)}(z_0; b) & \text{if } k \text{ is odd.} \end{cases}$$

Equation 11.4 is another analogue of the binomial formula. This theorem follows immediately from the expansion theorem 10.3, with the aid of (7.3) and (10.25). We note that since (11.4) is a relation between iterated integrals of the functions σ_i and τ_i, the equation holds independently of the expansion theorem and of the analyticity of the σ_i and τ_i.

We now define

$$(11.6)\qquad U_{Nn\nu}(x, y) = U_{Nn\nu}(0, 0; x, y).$$

From Theorem 11.2, taking the origin as b, we obtain from (11.1) and (11.6)

$$(11.7)\qquad U_{Nn\nu}(x_0, y_0; x, y) = \sum_{m=0}^{n} \sum_{\mu=0}^{2N-1} b_{Nn\nu m\mu} U_{Nm\mu}(x, y),$$

where $n = 0, 1, \cdots; \nu = 0, 1, \cdots, 2N-1$, and the b's are real constants. Hence, we have the

THEOREM 11.3. *Each solution $u = v(x, y)$ of the equation $L^N u = 0$, which is of class $C^{(2n)}$ in a domain containing the point z_0, admits the expansion*

$$(11.8)\qquad u(x, y) = \sum_{n=0}^{\infty} \{ \sum_{\nu=0}^{2N-1} \sum_{m=0}^{n} \sum_{\mu=0}^{2N-1} c_{\kappa\nu m\mu} U_{Nm\mu}(x, y) \},$$

where the c's are real constants and the series converges uniformly and absolutely in some neighborhood of z_0.

Finally, we observe that by (7.4), (7.5), and (7.10) to (7.13), the $U_{Nn\nu}$ are linear combinations with real constant coefficients of the functions given in (7.9), which are independent of N.

BROWN UNIVERSITY.

BIBLIOGRAPHY

1. L. Bers and A. Gelbart, "On a class of functions defined by partial differential equations," *Transactions of the American Mathematical Society*, vol. 56 (1944), pp. 67-93.
2. P. W. Ketchum, "Analytic functions of hypercomplex variables," *Transactions of the American Mathematical Society*, vol. 30 (1928), pp. 641-667.
3. G. Scheffers, "Verallgemeinerung der Grundlagen der gewohnlich complexen Funktionen," 1, 11, *Berichte uber die Verhandlungen der Sachsischen Akademie der Wissenschaft zu Leipzig, Mathematisch-physikalische Klasse*, vol. 45 (1893), pp. 828-848; vol. 46 (1894), pp. 120-134.
4. L. Sobrero, *Theorie der ebenen Elastizitat unter Benutzung eines Systems hyperkomplexer Zahlen*, Hamburger Mathematische Einzelschriften, Leipzig, 1934.
5. J. A. Ward, "A theory of analytic functions in linear associative algebras," *Duke Mathematical Journal*, vol. 7 (1940), pp. 233-248.

ON HIGHER NORMAL SPACES FOR V_m IN S_n.*

By ABRAHAM SCHWARTZ

C. Segre has proved that a real V_m in S_n whose points are all axial points is either contained in a geodesic V_{m+1} or has a second fundamental tensor of rank 1.[1] N. Coburn has studied the case of a real V_m in S_n whose points are all planar.[2] C. B. Allendoerfer has pointed out that if an open, simply connected domain of a Riemannian manifold imbedded within a Euclidean manifold is of "type" ≥ 3 at every point, then the second and higher normal spaces vanish.[3]

In this paper we consider a V_m in an S_n, and let the first normal space be of k dimensions at a particular point P. We prove first

THEOREM 1. *If the second normal space is of j dimensions, then*

$$j \leq \tfrac{1}{6}t(t+1)(t+2) + \tfrac{1}{2}u(u+1),$$

where t and u are integers determined by k in accordance with the following inequalities:

$$\tfrac{1}{2}t(t+1) \leq k < \tfrac{1}{2}(t+1)(t+2), \qquad u = k - \tfrac{1}{2}t(t+1).$$

Then we prove

THEOREM 2. *Either*

1) *The second normal space at the point P vanishes, or;*

2) *The three index curvature tensor, $\overset{1}{H}_{ab}{}^{\lambda}$, is of rank $\leq k$ with respect to the index a at P, or;*

3) *The V_m can be imbedded in a V_{m+r}, $0 < r < k$, in such a way that*

* Received April 22, 1946.

[1] Schouten and Struik, *Einfuhrung in die Neueren Methoden der Differentialgeometrie*, vol. 2, Groningen, p. 99.

[2] N. Coburn, "V_m in S_n with planar points ($m \geq 3$)" *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 774-783.

[3] C. B. Allendoerfer, "Rigidity for spaces of class greater than one," *American Journal of Mathematics*, vol. 61 (1939), p. 636.

$\overset{1*}{H}_{ab}{}^{\lambda}$, *the three index curvature tensor for the* V_m *in the* V_{m+r}, *is of rank* $\leq r$ *with respect to the index* a *at* P.[4]

Note that if the first normal space is everywhere of k dimensions and that if the second normal space vanishes everywhere, then the V_m can be imbedded in an S_{m+k}.

1. If k and j are the dimensionalities of the first and second normal spaces, one of the Ricci equations for a V_m in an S_n can be written in the form

$$(1.1) \qquad \overset{1}{H}_{ab}{}^{\nu}\,\overset{2}{H}_{c\nu}{}^{\lambda} = \overset{1}{H}_{cb}{}^{\nu}\,\overset{2}{H}_{a\nu}{}^{\lambda}, \qquad a, b, c = 1, \cdots, m;\ \nu, \lambda = 1, \cdots, n,$$

where the rank of $\overset{1}{H}_{ab}{}^{\nu}$ with respect to the index ν is k, and the rank of $\overset{2}{H}_{c\nu}{}^{\lambda}$ with respect to the index λ is j.[5] Since $\overset{1}{H}_{ab}{}^{\nu}$ is symmetric with respect to its first two indices, (1.1) says that the tensor

$$(1.2) \qquad J_{abc}{}^{\lambda} = \overset{1}{H}_{ab}{}^{\nu}\,\overset{2}{H}_{c\nu}{}^{\lambda},$$

which is of rank j with respect to the index λ, is symmetric with respect to its first three indices. Because of this symmetry we get the bound

$$(1.3) \qquad j \leq \tfrac{1}{6}m(m+1)(m+2)$$

immediately by considering the number of combinations of m things taken 3 at a time, repetitions allowed,[6] but this bound can be improved considerably.

Whenever

$$(1.4a) \qquad \overset{1}{H}_{ab}{}^{\nu} = x\overset{1}{H}_{cd}{}^{\nu} + y\overset{1}{H}_{ef}{}^{\nu} + \cdots + z\overset{1}{H}_{ij}{}^{\nu},$$

we have

$$(1.4b) \qquad J_{abg}{}^{\lambda} = \overset{1}{H}_{ab}{}^{\nu}\,\overset{2}{H}_{g\nu}{}^{\lambda} = xJ_{cdg}{}^{\lambda} + yJ_{efg}{}^{\lambda} + \cdots + zJ_{ijg}{}^{\lambda}.$$

One can select a set of k independent vectors of the first normal space from the set $\overset{1}{H}_{ab}{}^{\lambda}$, $a, b = 1, \cdots, m$. Let α be the number of these vectors which have 1 as one of their indices. By interchanging indices if necessary, we can be sure that no greater number of these vectors have some other integer as

[4] For the definition of the three index curvature tensor see Schouten and Struik, vol. 2, p. 80. For the definition of rank with respect to a particular index, see Schouten and Struik, vol. 1, p. 19.

[5] See Schouten and Struik, vol. 2, pp. 119, 122, and 124.

[6] A better bound than that given in Schouten and Struik, vol. 2, p. 116.

one of their indices. Let β be the number of the remaining k-α vectors which have 2 as one of their indices. Again we can assume that no greater number of the remaining vectors have some other integer as one of their indices. We repeat this process until the k vectors have been exhausted, and then we can list our k independent vectors as

$$
\begin{gathered}
\overset{1}{H}_{1a_1}{}^{\lambda}, \overset{1}{H}_{1a_2}{}^{\lambda}, \cdots, \overset{1}{H}_{1a_\alpha}{}^{\lambda} \\
\overset{1}{H}_{2b_1}{}^{\lambda}, \overset{1}{H}_{2b_2}{}^{\lambda}, \cdots, \overset{1}{H}_{2b_\beta}{}^{\lambda} \\
\cdots\cdots \\
\overset{1}{H}_{sn_1}{}^{\lambda}, \overset{1}{H}_{sn_2}{}^{\lambda}, \cdots, \overset{1}{H}_{sn_\nu}{}^{\lambda},
\end{gathered}
\tag{1.5}
$$

where

$$\alpha \geqq \beta \geqq \gamma \geqq \cdots \geqq \nu, \qquad \alpha + \beta + \gamma + \cdots + \nu = k. \tag{1.6}$$

Consider the vectors $J_{ab1}{}^{\lambda}$ of the second normal space. Because $\overset{1}{H}_{ab}{}^{\lambda}$ is a linear combination of the vectors in list (1.5), the vectors $J_{ab1}{}^{\lambda}$ are linear combinations of the vectors

$$J_{1a_{11}}{}^{\lambda}, J_{1a_{21}}{}^{\lambda}, \cdots, J_{1a_{\alpha 1}}{}^{\lambda}, J_{2b_{11}}{}^{\lambda}, \cdots, J_{2b_{\beta 1}}{}^{\lambda}, \cdots, J_{sn_{11}}{}^{\lambda}, \cdots, J_{sn_{\nu 1}}{}^{\lambda}. \tag{1.7}$$

Consider next the vectors $J_{ab2}{}^{\lambda}$. They are linear combinations of the vectors

$$J_{1a_{12}}{}^{\lambda}, J_{1a_{22}}{}^{\lambda}, \cdots, J_{1a_{\alpha 2}}{}^{\lambda} \tag{1.8}$$

and

$$J_{2b_{12}}{}^{\lambda}, J_{2b_{22}}{}^{\lambda}, \cdots, J_{2b_{\beta 2}}{}^{\lambda}, J_{3c_{12}}{}^{\lambda}, \cdots, J_{sn_{12}}{}^{\lambda}, \cdots, J_{sn_{\nu 2}}{}^{\lambda}. \tag{1.9}$$

Because $J_{1ac}{}^{\lambda} = J_{ac1}{}^{\lambda}$ the vectors of list (1.8) are dependent on those of list (1.7), but those of list (1.9) are perhaps independent. Similarly, when we consider the vectors $J_{ab3}{}^{\lambda}$ we will introduce the new vectors

$$J_{3c_{13}}{}^{\lambda}, \cdots, J_{3c_{\gamma 3}}{}^{\lambda}, \cdots, J_{sn_{13}}{}^{\lambda}, \cdots, J_{sn_{\nu 3}}{}^{\lambda}, \tag{1.10}$$

and finally, when we consider the vectors $J_{abs}{}^{\lambda}$ we will introduce the new vectors

$$J_{sn_{1s}}{}^{\lambda}, J_{sn_{2s}}{}^{\lambda}, \cdots, J_{sn_{\nu s}}{}^{\lambda}. \tag{1.11}$$

But any vector $J_{abc}{}^{\lambda}$ is linearly dependent on the vectors $J_{1a_1c}{}^{\lambda}, J_{1a_2c}{}^{\lambda}, \cdots, J_{1a_\alpha c}{}^{\lambda}, J_{2b_1c}{}^{\lambda}, \cdots, J_{sn_\nu c}{}^{\lambda}$, which are in turn dependent on the vectors of lists (1.7), (1.9), (1.10), (1.11) because of the J tensor symmetry. Thus, counting the vectors in (1.7), (1.9), (1.10), (1.11), we have

$$j \leqq 1\alpha + 2\beta + 3\gamma + \cdots + s\nu. \tag{1.12}$$

We shall show below that a coordinate system can always be found in which we shall have not only (1.6), but

$$\alpha > \beta > \gamma > \cdots > \nu. \tag{1.13}$$

Theorem 1 now follows from (1.12) and (1.13). For, to make the right hand side of (1.12) a maximum, one must choose

$$\begin{gathered}\alpha = t+1,\ \beta = t,\ \gamma = t-1, \cdots, \epsilon = t+2-u,\ \zeta = t+1-(u+1),\\ \eta = t+1-(u+2), \cdots, \nu = 1,\end{gathered} \tag{1.14}$$

where the integers t and u have been determined by k in the following way:

$$\tfrac{1}{2}t(t+1) \leq k < \tfrac{1}{2}(t+1)(t+2), \qquad u = k - \tfrac{1}{2}t(t+1). \tag{1.15}$$

The maximum value of the right hand side of (1.12) is then

$$\begin{aligned}&\sum_{n=1}^{u} n(t+2-n) + \sum_{n=u+1}^{t} n(t+1-n)\\ &\qquad = \tfrac{1}{6}t(t+1)(t+2) + \tfrac{1}{2}u(u+1).\end{aligned} \tag{1.16}$$

2. To complete the proof of Theorem 1 it is necessary to demonstrate (1.13).

LEMMA 1. *We can always assume that the index a_1 in the set* (1.5) *is* 1.

The lemma is obviously true if $\overset{1}{H}{}_{11}{}^{\lambda} \neq 0$. If $\overset{1}{H}{}_{11}{}^{\lambda} = 0$, we can assume that $\overset{1}{H}{}_{z1}{}^{\lambda} \neq 0$ and choose x so that the coordinate transformation

$$A_{a'}{}^{a} = \begin{cases} 1 \text{ if } a' = a \\ x \text{ if } a' = 1',\ a = z \\ 0 \text{ otherwise} \end{cases} \tag{2.1}$$

will carry us into a new coordinate system in which the lemma is satisfied.

LEMMA 2. *We can assume that the indices $b_1, b_2, \cdots, b_\beta$ in the set* (1.5) *are included within the indices $a_1, \cdots, a_\alpha$.*

For, suppose that $b_1, b_2, \cdots, b_i$ are not included in the set $a_1, a_2, \cdots, a_\alpha$. Then the coordinate transformation

$$A_{a'}{}^{a} = \begin{cases} 1 \text{ if } a' = a \\ x \text{ if } a' = 1',\ a = 2 \\ 0 \text{ otherwise} \end{cases} \tag{2.2}$$

will leave unchanged any vector $\overset{1}{H}_{ab}{}^{\lambda}$ for which neither a nor b is 1, and will replace $\overset{1}{H}_{11}{}^{\lambda}$ by $\overset{1}{H}_{1'1'}{}^{\lambda} = \overset{1}{H}_{11}{}^{\lambda} + 2x\overset{1}{H}_{12}{}^{\lambda} + x^2\overset{1}{H}_{22}{}^{\lambda}$, $\overset{1}{H}_{a1}{}^{\lambda}$ by $\overset{1}{H}_{a'1'}{}^{\lambda} = \overset{1}{H}_{a1}{}^{\lambda} + x\overset{1}{H}_{a2}{}^{\lambda}$, $a \neq 1$. The number x can be chosen so that there are $\alpha + i$ independent vectors in the set $\overset{1}{H}_{a_1'1'}{}^{\lambda}, \overset{1}{H}_{a_2'1'}{}^{\lambda}, \cdots, \overset{1}{H}_{a_\alpha'1'}{}^{\lambda}, \overset{1}{H}_{b_1'1'}{}^{\lambda}, \cdots, \overset{1}{H}_{b_i'1'}{}^{\lambda}$, and Lemma 2 holds in the new coordinate system.

These lemmas, together with the fact that $\overset{1}{H}_{12}{}^{\lambda} = \overset{1}{H}_{21}{}^{\lambda}$, make it possible to assert that $\alpha > \beta$. By proceeding in a similar fashion we can show that all of the inequalities of (1.13) hold.

3. The Ricci equation (1.1) can be rewritten in the form

$$\overset{p}{h}_{ab}\underset{p}{\overset{Q}{v}}_c = \overset{d}{h}_{cb}\underset{p}{\overset{Q}{v}}_a, \qquad p = 1, \cdots, k;\ Q = 1, \cdots, j;\ a, b, c = 1, \cdots, m,^{7} \tag{3.1}$$

the vanishing of the $\underset{p}{\overset{Q}{v}}_c$ being a necessary and sufficient condition that the second and higher normal spaces vanish. Let the index Q be fixed and write $\underset{p}{\overset{Q}{v}}_c \equiv \underset{p}{v}_c$. If the rank of the matrix

$$\left\| \begin{matrix} \underset{1}{v_1} & \underset{2}{v_1} & \cdots & \underset{k}{v_1} \\ \underset{1}{v_2} & \underset{2}{v_2} & \cdots & \underset{k}{v_2} \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \underset{1}{v_m} & \underset{2}{v_m} & \cdots & \underset{k}{v_m} \end{matrix} \right\| \tag{3.2}$$

is 0 no matter what index Q has been chosen, then $j = 0$ and the second normal space vanishes. If for some choice of Q the rank of (3.2) is r, $0 < r \leq k$, then one can find $m - r$ linearly independent vectors $\underset{i}{\lambda}{}^c$, $i = 1, \cdots, m - r$, with the property

$$\underset{i}{\lambda}{}^c \underset{p}{v}_c = 0, \qquad i = 1, \cdots, m - r;\ p = 1, \cdots, k, \tag{3.3}$$

and it would follow from (3.1) that

$$\underset{i}{\lambda}{}^c \underset{p}{v}_a \overset{p}{h}_{cb} = 0. \tag{3.4}$$

---

[7] Apply equations (13.47) and (13.45), p. 126, Schouten and Struik, vol. 2, to (1.1) of this paper. Equations (13.53c), p. 128, Schouten and Struik, vol. 2 are

But we can always replace the second fundamental tensors $\overset{p}{h}_{ab}$ by a new set $\overset{p}{l}_{ab}$ derived from the old by using an orthogonal transformation on the index p. Since the matrix (3.2) is of rank r, it is possible to choose the index a in (3.4) in r essentially different ways and to find a new set of second fundamental forms, $\overset{p}{l}_{ab}$, such that

$$(3.5) \qquad \underset{i}{\lambda}^c \overset{\alpha}{l}_{cb} = 0, \qquad \alpha = 1, \cdots, r;\ i = 1, \cdots, m - r.$$

If $r = k$, (3.5) becomes

$$(3.6) \qquad \underset{i}{\lambda}^a \overset{p}{l}_{ab} = 0, \qquad p = 1, \cdots, k;\ i = 1, \cdots, m - k.$$

Since

$$(3.7) \qquad \overset{1}{H}_{ab}{}^{\nu} = \overset{p}{l}_{ab} \underset{p}{i}^{\nu},$$

we have

$$(3.8) \qquad \underset{i}{\lambda}^a \overset{1}{H}_{ab}{}^{\nu} = 0, \qquad i = 1, \cdots, m - k.$$

Hence, in this case, the rank of $\overset{1}{H}_{ab}{}^{\nu}$ with respect to the index a is at most k.

If $r < k$, the local R_m and the r unit vectors $\underset{\alpha}{i}^{\nu}$ of the first normal space which correspond to the r second fundamental forms $\overset{\alpha}{l}_{ab}$ mentioned in (3.5) will determine an $(m + r)$-dimensional direction at the point P. A V_{m+r} which contains the V_m and this $(m + r)$-dimensional direction at the point P can then be found, and if we let $\overset{1*}{H}_{ab}{}^{\nu}$ be the three index curvature tensor of the V_m with respect to this V_{m+r}, we can write

$$(3.9) \qquad \overset{1*}{H}_{ab}{}^{\nu} = \overset{c}{l}_{ab} \underset{a}{i}^{\nu}.^{8}$$

(3.5) can then be written as

$$(3.10) \qquad \underset{i}{\lambda}^a \overset{1*}{H}_{ab}{}^{\nu} = 0, \qquad i = 1, \cdots, m - r,$$

and hence the rank of $\overset{1*}{H}_{ab}{}^{\nu}$ with respect to the index a is at most r in this case. Theorem 2 has now been demonstrated.

---

correct only for the cases $y \geq 3$. Equation (3.1) of this paper is the correct statement for the case $y = 2$ if the V_n is an S_n.

[8] See Schouten and Struik, vol. 2, p. 91.

4. Allendoerfer has shown that the second normal space vanishes at a particular point if the "type" there is ≥ 3. When the type is ≥ 3, one can find a coordinate system in which the matrix

$$(4.1)\qquad \begin{Vmatrix} \overset{1}{l}_{11} & \overset{1}{l}_{21} & \overset{1}{l}_{31} & \overset{2}{l}_{11} & \overset{2}{l}_{21} & \overset{2}{l}_{31} & \cdots & \overset{k}{l}_{11} & \overset{k}{l}_{21} & \overset{k}{l}_{31} \\ \overset{1}{l}_{12} & \overset{1}{l}_{22} & \overset{1}{l}_{32} & \overset{2}{l}_{12} & \overset{2}{l}_{22} & \overset{2}{l}_{32} & \cdots & \overset{k}{l}_{12} & \overset{k}{l}_{22} & \overset{k}{l}_{32} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \overset{1}{l}_{1m} & \overset{1}{l}_{2m} & \overset{1}{l}_{3m} & \overset{2}{l}_{1m} & \overset{2}{l}_{2m} & \overset{2}{l}_{3m} & \cdots & \overset{k}{l}_{1m} & \overset{k}{l}_{2m} & \overset{k}{l}_{3m} \end{Vmatrix}$$

is of rank $3k$.[9] It would not now be possible to find a set of vectors $\underset{i}{\lambda}{}^{c}$, $i = 1, \cdots, m - r$, $r \leq k$ which satisfies (3.5), for if such vectors existed, (3.5) would say directly that the matrix

$$(4.2)\qquad \begin{Vmatrix} \overset{1}{l}_{11} & \overset{1}{l}_{21} & \overset{1}{l}_{31} & \overset{2}{l}_{11} & \overset{2}{l}_{21} & \overset{2}{l}_{31} & \cdots & \overset{r}{l}_{11} & \overset{r}{l}_{21} & \overset{r}{l}_{31} \\ \overset{1}{l}_{12} & \overset{1}{l}_{22} & \overset{1}{l}_{32} & \overset{2}{l}_{12} & \overset{2}{l}_{22} & \overset{2}{l}_{32} & \cdots & \overset{r}{l}_{12} & \overset{r}{l}_{22} & \overset{r}{l}_{32} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot \\ \overset{1}{l}_{1m} & \overset{1}{l}_{2m} & \overset{1}{l}_{3m} & \overset{2}{l}_{1m} & \overset{2}{l}_{2m} & \overset{2}{l}_{3m} & \cdots & \overset{r}{l}_{1m} & \overset{r}{l}_{2m} & \overset{r}{l}_{3m} \end{Vmatrix}$$

was of rank $\leq r$, and hence that (4.1) was of rank $< 3k$. Thus, whenever Allendoerfer's theorem predicts the vanishing of the second normal space, so does Theorem 2.

On the other hand, there are cases where Theorem 2 predicts the vanishing of the second normal space at a particular point while Allendoerfer's theorem does not. The simplest such cases arise when $k = 2$, $m = 3$. In these cases, the "type" would be ≤ 1, and hence Allendoerfer's Theorem would make no prediction. But one can easily find two three-rowed symmetric matrices for which vectors $\underset{i}{\lambda}{}^{c}$, $i = 1, \cdots, 3 - r$, $r \leq 2$, which satisfy (3.5) cannot be found, and in these cases Theorem 2 would predict the vanishing of the second normal space.

THE PENNSYLVANIA STATE COLLEGE.

---

[9] C. B. Allendoerfer, "Rigidity for spaces of class greater than one," *American Journal of Mathematics*, vol. 61 (1939), pp. 635 and 636.

CONCERNING CONTINUA IRREDUCIBLE ABOUT n POINTS.*

By R. H. Sorgenfrey.

It is the purpose of this paper to characterize, in terms of combining properties of subcontinua, continua in metric spaces which are irreducible about n points.[1] R. L. Moore has proved [2] that if a plane atriodic continuum contains no subcontinuum which separates the plane, then it is irreducible about some two points. The present author later generalized this result by showing [3] that if a unicoherent continuum fails to be a triod, then it is irreducible about some two points. The conditions of the hypothesis of this theorem, although sufficient for irreducibility, are not necessary, as simple examples exist of non-unicoherent continua which are irreducible. In this paper a condition is given which is both necessary and sufficient for irreducibility about n points.

A finite number of subcontinua of a continuum M will be said to form a *proper decomposition* of M provided that their sum is M and that no one of them is a subset of the sum of the others.

Theorem. *In order that the continuum M be irreducible about some n points, $n \geqq 2$, it is necessary and sufficient that for every proper decomposition of M into $n+1$ continua, the sum of some n of these fails to be connected.*

The necessity of the condition is obvious. The proof of its sufficiency will be shortened if the condition is formulated as follows: The continuum M will be said to have *property B_n* provided that for every decomposition of M into $n+1$ continua, the sum of some n of these fails to be connected. A proper decomposition of M into $n+1$ continua, the sum of each n of which is connected, will be called a *B_n contradicting decomposition* of M (abbre-

* Received May 14, 1946. Presented to the American Mathematical Society, November 24, 1945.

[1] The term "irreducible" as here used pertains to subcontinua: A continuum M is irreducible about the n points $P_1, P_2, \cdots, P_n$ provided that the only subcontinuum of M containing all these points is M itself. A characterization of connected spaces irreducible about n points relative to connected subsets was obtained in 1931 by E. Čech ("Sur les ensembles connexes irreducibles entre n points," *Casopis*, vol. 61, pp. 109-129).

[2] "Concerning compact continua which contain no subcontinuum which separates the plane," *Proceedings of the National Academy of Sciences*, vol. 20 (1934), pp. 41-45.

[3] "Concerning triodic continua," *American Journal of Mathematics*, vol. 66 (1944), pp. 439-460.

viated B_n c. d.). Thus a B_1 c. d. is simply a decomposition of M into two proper subcontinua. The proof of the sufficiency will be by induction. Two lemmas will first be established.

LEMMA 1. *Suppose that the continuum M has property B_n ($n \geqq 2$) and that the continua $H_1, H_2, \cdots, H_n$ form a B_{n-1} c. d. of M. Denote by S_r the sum of all the continua $H_1, H_2 \cdots, H_n$ except H_r, and let $D_r = M - S_r$. Then* (a) *D_r is connected,* (b) *there does not exist a decomposition of $\bar{D}_r$ into two proper subcontinua each of which intersects S_r, and* (c) *if D_r is indecomposable, not every composant*[4] *of $\bar{D}_r$ intersects S_r.*

Proof. The set $\bar{D}_r$ cannot have infinitely many components for if it did it would be possible to construct a B_n c. d. of M into continua[5] of the form $S_r + U_j$ ($j = 1, \cdots, n+1$) where the U_j's are mutually separated sets whose sum is D_r and whose components are components of D_r. Hence each component of D_r is open in M. Suppose that D_r has more than one component and denote the closures of two of them by E and F. Let R be the sum of all components of $H_r - (E + F)$ which have limit points in E and let T be the sum of all those which have limit points in F. Then $E + \bar{R}$ and $F + \bar{T}$ are distinct continua whose sum is H_r. But these two continua, together with all the continua $H_1, \cdots, H_n$, except H_r, form a B_n c. d. of M. This establishes (a).

To prove (b) suppose that $\bar{D}_r$ is the sum of the proper subcontinua E and F, each of which intersects S_r. Since neither E nor F contains D_r, each of them contains an open subset of D_r not intersecting the other. Hence an argument identical with the one given in the paragraph above will lead to a contradiction.

To prove (c) suppose that $\bar{D}_r$ is indecomposable and that each of its composants intersects S_r. Let G be the upper semi-continuous collection whose elements are S_r and the individual points of D_r, let M' be the hyperspace of the decomposition G, and let f be the associated transformation. Then f is continuous and monotone. The continuum M' is not indecomposable since each two of its points lie on a proper subcontinuum of M'. It is therefore the sum of two proper subcontinua E' and F'. Let $f^{-1}(E') = E$ and $f^{-1}(F') = F$. Then E and F are proper subcontinua of M each of which

[4] By a *composant* of an indecomposable continuum is meant a subset which is maximal with respect to the properties of being continuumwise connected and proper.

[5] The fact that $S_r + U_j$ is a continuum follows from a modification of a theorem of W. A. Wilson: *If K is a closed subset of the continuum M, then every component of $M - K$ has a limit point in K.* This theorem will be used repeatedly in the proof.

contains S_r. Since $M - E$ can have only a finite number of components, each of them is an open subset of $\bar{D}_r$. Let U be one of these components. Then $\bar{U}$ is a proper subcontinuum of $\bar{D}_r$ which contains an open subset of $\bar{D}_r$ which, since $\bar{D}_r$ is indecomposable, is impossible.

LEMMA 2. *If M has property B_2 and H and K are subcontinua of M such that $M - H$ and $M - K$ are connected and $H \cdot (M - K)$ and $K \cdot (M - H)$ contain open subsets of M, then any subcontinuum L of M which intersects both H and K contains $M - (H + K)$.*

Proof. Suppose that the lemma is not true. The set $U = M - (H + K + L)$ is connected by Lemma 1(a) (for $n = 2$, taking $H_1 = H + K + L$ and $H_2 = \overline{M - H}$). The continuum $\bar{U}$ cannot intersect both $H + L$ and $K + L$ for, if it did, these three continua would form a B_2 c. d. of M. Hence $\bar{U}$ fails to intersect $H + L$, say. But then $M - K$, which intersects both U and $H + L$ and is a subset of their sum, would fail to be connected, contrary to hypothesis.

Proof of the sufficiency of the condition. The proof will first be given for $n = 2$, that is, it will be shown that if the continuum M has property B_2, it is irreducible between some two points. Since every indecomposable continuum is irreducible about some two points, it will be assumed that M is the sum of two proper subcontinua H and K. Let $M - K = U$ and $M - H = V$. It will now be shown that there exists a point X of U such that any subcontinuum of M containing X and a point of K contains U. Suppose that no such point exists.

By Lemma 1(a) U is connected. Denote it by U_0 and $\bar{U} \cdot K$ by N_0. There exists a countable set $C = P_1 + P_2 + \cdots$ of U_0 such that $\bar{C} = \bar{U}_0$. Denote P_1 by X_1. There exists a subcontinuum H_1 of $\bar{U}_0$ irreducible from X_1 to N_0. Let $N_1 = N_0 + H_1$. If $U_0 - N_1$ is non-vacuous, it is, by Lemma 1(a), connected; denote it by U_1. Let n_2 be the smallest integer n such that P_n belongs to U_1, and denote P_{n_2} by X_2. There exists a subcontinuum H_2 of $\bar{U}_1$ irreducible from X_2 to N_1. Let $N_2 = N_1 + H_2$. This process may be continued unless for some integer r, $U_0 - N_r$ is vacuous. Suppose this to be the case and denote X_r by X. Let L be a subcontinuum of M containing X and a point of K. Since H_r is irreducible from X to N_{r-1}, it is a subset of L, and since, by Lemma 2, L also contains $M - (K + H_r)$, it contains U. Therefore it follows from the initial supposition that the process described above is an infinite one.

Suppose first that for each i there exists a j such that $N_i \cdot \bar{U}_j = 0$. The

set $X_1 + X_2 \cdots$ has a limit point X. Under the initial supposition there exists a subcontinuum L of M containing X and a point of K but not all of U. If, for each i, L contained $N_i - N_0$, it would contain C. But since $\bar{C} = \bar{U}$, L would then contain U. Hence for some i, L fails to contain $N_i - N_0$. There exists a j such that $N_i \cdot \bar{U}_j = 0$. The point X belongs to $\bar{U}_j$ since all but a finite number of the points $X_1, X_2, \cdots$ belong to U_j and hence, by Lemma 2, L contains $M - (K + \bar{U}_j)$ which contains $N_i - N_0$, a contradiction.

It therefore follows that for some i, $\bar{U}_i \cdot U_j \neq 0$ for each j. Suppose that the continuum $\bar{U}_i$ is the sum of two proper subcontinua E and F. By Lemma 1(b) one of these, say F, fails to intersect N_i. For some k, $F \cdot N_k \neq 0$. But N_k does not contain E; for if it did, U_k would be a subset of F and hence $N_i \cdot \bar{U}_k = 0$. But then E, $F + N_k$, and $K + N_k$ form a B_2 c. d. of M. It follows from this contradiction that $\bar{U}_i$ is indecomposable. By Lemma 1(c) some composant of $\bar{U}_i$ fails to intersect $K + N_i$. Let X be a point of such a composant and let L be a subcontinuum of M containing X and a point of K. Then L contains $\bar{U}_i$ and, by Lemma 2, also contains $M - (K + \bar{U}_i)$ and hence U, a contradiction of the original assumption.

It follows that there exist points X of U and Y of V such that any subcontinuum of M containing both X and Y contains $U + V$. But, by Lemma 2, such a continuum also contains $M - (\bar{U} + \bar{V})$ and must therefore be M itself. This completes the proof for the case $n = 2$.

Assume now that the sufficiency has been established for $n = k - 1$ and assume that it is not true for $n = k$. Let M be a continuum having property B_k which is not irreducible about any k points. Then it does not have property B_{k-1} since any continuum irreducible about $k - 1$ points is irreducible about k points. There therefore exist continua $H_1, H_2, \cdots, H_k$ which form a B_{k-1} c. d. of M.

By Lemma 1(a) (using the notation of that lemma) each of the sets D_r is connected. It follows from Lemma 1(b) that $\bar{D}_r$ has property B_2, for if K_1, K_2, and K_3 form a B_2 c. d. of $\bar{D}_r$, at least one of them, say K_1, intersects S_r and then $K_1 + K_2$ and $K_1 + K_3$ would form a non-allowable decomposition of $\bar{D}_r$. Hence, by the sufficiency for $n = 2$, $\bar{D}_r$ is irreducible between some two points.

If $\bar{D}_r$ is indecomposable one of its composants fails to intersect S_r by Lemma 1(c). Hence it is irreducible between a point X_r of such a composant and S_r, and every subcontinuum of M containing X_r and a point of S_r contains $\bar{D}_r$. If $\bar{D}_r$ is not indecomposable, let E and F be proper subcontinua whose sum is $\bar{D}_r$. By Lemma 1(b) one of these, say F, fails to intersect S_r. Then $\bar{D}_r$ is irreducible between some point X_r of $\bar{D}_r - E$ and a point of E.

Let L be a subcontinuum of M intersecting both X_r and S_r. Let L' be the closure of the component of $L - L \cdot S_r$ which contains X_r. Because of the irreducibility of $\bar{D}_r$, L' contains $\bar{D}_r - E$. It must also contain E for otherwise the decomposition of $\bar{D}_r$ into the continua E and $L' + \overline{\bar{D}_r - E}$ would contradict Lemma 1(b). In either case, therefore, D_r contains a point X_r such that any subcontinuum of M intersecting both X_r and S_r contains D_r.

By the initial supposition there exists a proper subcontinuum L of M which contains $X_1 + X_2 + \cdots + X_k$. By the preceding paragraph L contains $D_1 + \cdots + D_k$. Since the number of components of $M - L$ is finite, each of them is open in M. Denote one of them by D_{k+1}, and let $S_{k+1} = M - D_{k+1}$. For each r $(r = 1, \cdots, k+1)$, let $C_j{}^r$ be the closure of the sum of all components of $S_r - (\bar{D}_1 + \cdots + \bar{D}_{r-1} + \bar{D}_{r+1} + \cdots + \bar{D}_{k+1})$ which have limit points in $\bar{D}_j$ $(j = 1, \cdots, r-1, r+1, \cdots, k+1)$. Then $K_j = \bar{D}_j + C_j{}^1 + \cdots + C_j{}^{j-1} + C_j{}^{j+1} + \cdots + C_j{}^{k+1}$ $(j = 1, \cdots, k+1)$ is a continuum not intersecting $D_1 + \cdots + D_{j-1} + D_{j+1} + \cdots + D_{k+1}$. Moreover $K_1 + \cdots + K_{r-1} + K_{r+1} + \cdots + K_{k+1} = S_r$ $(r = 1, \cdots, k+1)$ and is therefore connected. Hence the continua $K_1, \cdots, K_{k+1}$ form a B_k c. d. of M. This contradiction justifies the induction process and therefore establishes the theorem.

UNIVERSITY OF CALIFORNIA,
LOS ANGELES.

THE COEFFICIENT OF VISCOUS TRACTION.* [1]

By M. Reiner.

1. In the first paper the rheological equation of a special case of the generalized Newtonian liquid—the Reynolds Liquid—was derived. While in the generalized Newtonian liquid the tensors of stress and velocity-strain are connected by means of three scalar functions F_0, F_1, F_2 (or $\mathfrak{F}_0, \mathfrak{F}_1, \mathfrak{F}_2$) of *all three* tensor invariants, in the Reynolds liquid these functions are independent of the third invariant and subjected to certain other limitations. The thus specialized functions allow of physical interpretations through "rheological coefficients," viz. the coefficient of dilatancy δ, and the viscosities η_ν and η (or fluidities ϕ_ν and ϕ). We found

$$F_0(0, I'_2, 0)/I'_2 = \delta;\ F_0(e, 0, 0)/3e = \eta_\nu;\ F_1(0, I'_2, 0)/2 = \eta;\ F_2(0, I'_2, 0) = 0 \tag{78}$$ [2]

or

$$\mathfrak{F}_0(0, T'_2, 0) = 0;\ 3\mathfrak{F}_0(T, 0, 0)/T = \phi_\nu;\ 2\mathfrak{F}_1(0, T'_2, 0) = \phi;\ \mathfrak{F}_2(0, T'_2, 0) = 0 \tag{79}$$ [2]

where

$$\phi_\nu = 1/\eta_\nu;\ \phi = 1/\eta \tag{80}$$

from which we see that in certain cases the generalized Newtonian Liquid will behave not differently from the Reynolds Liquid. These cases cover the standard tests of viscometry in the tube- and rotation viscometers from which the rheological equation of the Reynolds liquid was derived in the first paper. In other words: If no other tests were available, the existence of any viscous liquid more general than the Reynolds liquid could not be detected. There exists, however, another basic test with which we did not deal in the first paper and which yields additional information as will be shown.

---

* Received October 10, 1945.

[1] This paper forms a continuation of my paper "A mathematical theory of dilatancy" in this Journal, vol. 67 (1945), pp. 350-362, which will be quoted as "the first paper."

[2] By a slip of the pen in the first paper η_ν was written when the intention was to write $3\eta_\nu$ and ϕ_ν was written for $\phi_\nu/3$. This has now been corrected. The present notation makes the analogy of η_ν with the bulk modulus (κ) perfect and is in agreement with the notation used by others.

2. For highly viscous substances to the order of 10^6 to 10^{10} poises such as pitch, shoemaker's wax and others, the standard "shear" methods of viscometry cannot be used. Trouton formed such substances into cylindrical rods which he subjected to pull and observed the rate of elongation. Let T_{zz} be the tractional force per unit area of cross section and e_{zz} the rate of elongation per unit length, then Trouton defined a coefficient of Viscous Traction λ^* by the equation

$$T_{zz} = \lambda^* e_{zz}. \tag{81}$$

The coefficient λ^* is the viscous analogue of Young's modulus and there must be some relation between λ^* and η similar to the relation between Young's modulus and the modulus of rigidity. Trouton attempted to show by reasoning and experiment that $\lambda^*/\eta = 3$, but his not very numerous experiments actually gave an average of $\lambda^*/\eta = 3.16$ and in his reasoning he overlooked that not one of his materials was a *simple* Newtonian liquid. We shall therefore take up the problem as one of the rheological behavior of a *generalized* Newtonian liquid.

3. We have to distinguish in Trouton's experiment two stages. The stress due to the pull T_{zz} has an isotropic component $T_{zz}/3 \cdot \delta_s{}^r$. When the pull is applied, there will be an *initial stage* which starts with an accelerated and ends with a retarded movement of the particles, with between—generally—pulsations. During this initial stage the material expands, the measure of the cubical dilatation at every moment being 3ϵ. This cubical dilatation produces an elastic reaction (hydrostatic tension) $3\kappa\epsilon$. It is accompanied by a viscous resistance $3\eta_v e$ due to volume viscosity. When ϵ has so much increased that $3\kappa\epsilon = T_{zz}/3$, the elastic reaction balances the isotropic component of the pull T_{zz}. The cubical dilatation then ceases to increase and e ultimately vanishes. With this the second stage sets in, in which the movement is steady.

The experiment is arranged in the same way as experiment (iii) of the first paper:—T_{zz} is kept constant and observations start after the steady state of continued elongation at constant rate has been reached. Let λ be the coefficient of viscous traction in this state, then $\lambda = \lambda(0, I'_2, I'_3)$. We are interested in the rheologics of the second stage of steady flow, but it will be profitable to include in the considerations the first stage also.

4. If we neglect the stresses due to the weight of the cylinder and to the inertia of its parts, in short if we neglect in the stress equations all terms in which the density ρ appears as a factor, a homogeneous stress distribution is possible and viscous traction may be defined by

$$T^*{}_s{}^r = T_{zz} \begin{Vmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{Vmatrix}. \qquad (82)$$

It will be convenient to introduce an axial unit tensor

$$\gamma_s{}^r = \begin{Vmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{Vmatrix}, \qquad (83)$$

the first invariant of which is 1/3 and whose deviator is accordingly

$$\gamma'{}_s{}^r = \gamma_s{}^r - 1/3 \cdot \delta_s{}^r = 1/3 \begin{Vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{Vmatrix} \qquad (84)$$

having the invariants

$$I'_{\gamma 2} = -1/3; \; I'_{\gamma 3} = 2/27. \qquad (85)$$

From (84), since $\gamma_a{}^r \gamma_s{}^a = \gamma_s{}^r$,

$$\gamma'{}_a{}^r \gamma'{}_s{}^a = 1/3 \cdot \gamma_s{}^r + 1/9 \cdot \delta_s{}^r = 1/3 \cdot \gamma'{}_s{}^r + 2/3 \cdot \delta_s{}^r. \qquad (86)$$

We now write (82)

$$T^*{}_s{}^r = T_{zz} \gamma_s{}^r \qquad (87)$$

and have, considering (8),

$$T = T_{zz}/3 - 3\kappa\epsilon; \; T'{}_s{}^r = T_{zz} \gamma'{}_s{}^r \qquad (88)$$

from which

$$T'{}_a{}^r T'{}_s{}^a = T^2{}_{zz}(1/3 \cdot \gamma'{}_s{}^r + 2/9 \cdot \delta_s{}^r), \; T'_2 = -T^2{}_{zz}/3; \; T'_3 = 2T^3{}_{zz}/27. \quad (89)$$

This gives (compare (49))

$$e'{}_s{}^r = T_{zz}(\mathfrak{F}_1 + \mathfrak{F}_2 T_{zz}/3) \gamma'{}_s{}^r. \qquad (90)$$

The coefficient of viscous traction as defined by (81) would therefore be

$$(\lambda^*)^{-1} = e/T_{zz} + 2/3 \cdot (\mathfrak{F}_1 + \mathfrak{F}_2 T_{zz}/3) \qquad (91)$$

which makes (90)

$$e'_s{}^r = 3/2 \cdot (T_{zz}/\lambda^* - e)\gamma'_s{}^r. \tag{92}$$

In the second stage, when e vanishes, these expressions are reduced to

$$(\lambda)^{-1} = 2/3 \cdot (\mathfrak{F}_1 + \mathfrak{F}_2 T_{zz}/3)\,;\ e'_s{}^r = 3/2 \cdot T_{zz}/\lambda \cdot \gamma'_s{}^r = 3/2 e_{zz}\gamma'_s{}^r. \tag{93}$$

5. Before elaborating these general equations, it will be useful to examine the conditions prevailing in the first approximation of the viscous liquid, namely, the simple Newtonian liquid, the rheological equation of which is

$$T^*{}_s{}^r = 3(\kappa\epsilon + \eta_\nu e)\delta_s{}^r + 2\eta e'_s{}^r. \tag{94}$$

From this

$$T^* = 3(\kappa\epsilon + \eta_\nu e). \tag{95}$$

On the other hand from (82)

$$T^* = T_{zz}/3 \tag{96}$$

and therefore

$$e = (T_{zz} - 9\kappa\epsilon)/9\eta_\nu. \tag{97}$$

Comparison of (94) with (41) gives

$$F_0(I') = 3\eta_\nu e\,;\ F_1(I') = 2\eta\,;\ F_2(I') = 0. \tag{98}$$

These expressions are of the same form as (78), but it should be noted that in the present case no restrictions have been placed on the arguments of the functions F nor on the coefficients η and η_ν, which need not be constants.

Introducing F_1 and F_2 from (98) into (53), we find

$$\mathfrak{F}_1 = 1/2 \cdot \eta\,;\ \mathfrak{F}_2 = 0 \tag{99}$$

which makes (91)

$$(\lambda^*)^{-1} = 1/3 \cdot \eta + e/T_{zz} \tag{100}$$

and considering (97)

$$\lambda^* = \frac{9\eta \dfrac{\eta_\nu}{1 - 9\kappa\epsilon/T_{zz}}}{\eta + 3\dfrac{\eta_\nu}{1 - 9\kappa\epsilon/T_{zz}}} \tag{101}$$

while, when e vanishes,

$$\lambda = 3\eta. \tag{102}$$

(101) is the viscous analogy to the classical relation between Young's modulus

(E), the modulus of rigidity (μ) and the bulk modulus (κ), viz. $E = 9\mu\kappa/(\mu + 3\kappa)$. It should, however, be noted that the analogy is *not exact*. While $E = 3\mu$ for $\kappa = \infty$ and *only* in this case,[3] the situation is very different with regard to λ^*, as will presently be discussed.

In interpreting (101), one should keep in mind that $1 - 9\kappa\epsilon/T_{zz} \geq 0$, where the sign of inequality is valid in the first, the sign of equality in the second, stage of Trouton's experiment. For the volume viscosity we have $0 \leq \eta_v \leq \infty$. It is clear that the volume viscosity cannot be negative as "otherwise the more alternate expansion and compression, alike in all directions, of a fluid, instead of demanding the exertion of work upon it, would cause it to give work out" (Stokes). As can be seen, in the first stage, if η_v vanishes, λ^* also vanishes. On the other hand $\lambda/\eta = 3$ in both stages if $\eta_v = \infty$ and also in the second stage *whatever the magnitude of* η_v. Because η_v cannot be negative, λ^*/η cannot exceed the value 3. Therefore $0 \leq \lambda^* \leq 3$, while $\lambda = 3\eta$.

These results are of interest in connection with the rheological equation of the classical viscous liquid or what may be called the Stokes Liquid, for which $T^*{}_s{}^r = -(p + 2\eta e)\delta_s{}^r + 2\eta e_s{}^r$. This was derived by Stokes assuming $\eta_v = 0$, but in the first stage of the Trouton experiment a vanishing volume viscosity would mean vanishing viscous resistance against extension of a liquid cylinder, no matter how high the ordinary viscosity η of the liquid—a result at variance with our ideas of viscous flow. Tisza has recently drawn attention to the following quotation from Stokes: . . . "of course we may at once put $\eta_v = 0$ if we assume that in the case of a uniform motion of dilatation the pressure at any instant depends only on the actual density and temperature at that instant and not at the rate at which the former changes with the time. In most cases in which it would be interesting to apply the theory of friction of fluids the density of the fluid is either constant or may without sensible error be regarded as constant, or else changes slowly with the time. In the first two cases the results would be the same and in the third nearly the same whether η_v were equal to zero or not. Consequently, if theory and experiments should in such cases agree, the experiments must not be regarded as confirming that part of the theory which relates to supposing η_v to be equal to zero." Examination of (101) shows that Stokes was mistaken in equalizing the influence of either $e = 0$ or $\eta_v = 0$ on experimental results. We see that the same result follows from either $e = 0$ or $\eta_v = \infty$ and *not* $\eta_v = 0$.

It therefore appears that there does not exist a *real* viscous liquid for

---

[3] Excepting the degenerate case $\mu = 0$.

which the Stokes liquid can serve as a suitable approximation. The first approximation for a viscous liquid is the simple Newtonian liquid possessing *two* constants. This liquid behaves in the more primitive way of the Stokes liquid when $e = 0$ and the Stokes liquid does not represent a *type* of liquid, but rather a rudimentary condition of another type (the simple Newtonian). The Stokes-Navier differential equations should therefore be used only for $e = 0$. The equation $e = 0$ is called the equation of continuity of the incompressible liquid, but it should be noted that while it is true that $e = 0$ is valid for every kind of flow of an incompressible liquid for which $\epsilon = 0$, it also applies to *certain* kinds of flow of *compressible* liquids, e. g., to laminar flow where $\epsilon = 0$, while the liquid is *not incompressible,* and also to such kinds of flow, as in the second stage of Trouton's experiment, where $\epsilon \neq 0$.

6. It is thought that in view of the standard treatment in most text-books, this question of the legitimacy of the concept of the Stokes liquid and therefore also of the Stokes-Navier differential equations is important enough to warrant some further discussion from another angle. Touton's experiment is the viscous analogy to the Young's modulus experiment dealt with by Murnaghan. We may start, as he does, from the kinematics of the case and assume

$$\dot{u} = qx; \ \dot{v} = qy; \ \dot{w} = rz. \qquad (103)^{4}$$

If we introduce these values into the Stokes-Navier differential equations, we find that the three components of the acceleration a_x, a_y, a_z must vanish or, alternatively, that $\rho a_x, \rho a_y, \rho a_z$ should be negligible. If this is the case, the stress will be homogeneous, so that the assumption (103) corresponds with the assumption (82).

From (103) we find

$$e_s{}^r = \left\| \begin{matrix} q & 0 & 0 \\ 0 & q & 0 \\ 0 & 0 & r \end{matrix} \right\| ; \ e = (2q + r)/3; \ e'_s{}^r = (r - q)\gamma'_s{}^r. \qquad (104)$$

For the simple Newtonian liquid

$$T^*{}_s{}^r = [-p + \eta\nu(2q + r)]\delta_s{}^r + 2\eta(r - q)\gamma'_s{}^r \qquad (105)$$

The numbers q and r must be such that the expressions on the right side of (105) and (82) can be equated.

This requires

[4] I have replaced Murnaghan's p by q, having disposed of p before.

$$q = (3p + T_{zz})/9\eta_v - T_{zz}/6\eta; \quad r = (3p + T_{zz})/9\eta_v + T_{zz}/3\eta$$
$$e = (3p + T_{zz})/9\eta_v; \quad e' = T_{zz}/2\eta \cdot \gamma'_s{}^r; \tag{106}$$
$$q/r = [2\eta(3p + T_{zz}) - 3\eta_v T_{zz}] / [2\eta(3p + T_{zz}) + 6\eta_v T_{zz}]. \tag{107}$$

The quotient q/r is the viscous analogy to Poisson's ratio. If we assume vanishing η_v, we find $q = \infty$, $r = \infty$, $e = \infty$, $q/r = 1$. The expansion of the material in the first stage would be purely elastic, instantaneous and be completed in no time. This is what would be expected in the absence of the damping influence of a volume viscosity. If, on the other hand, η_v would be so great that it may be put $= \infty$, we find $q = -T_{zz}/6\eta$; $r = T_{zz}/3\eta$; $e = 0$; $q/r = -\frac{1}{2}$ and the expansion would be infinitely delayed so that the second stage would never be reached. It may be mentioned that Tisza deduces from observations on supersonic absorption in certain liquids for η_v/η a value of 2000. This, for the purpose of our calculations, is practically infinite.

Applying (104) to the Stokes liquid, we get

$$p = -2\eta(r - q)/3; \; p + T_{zz} = 4\eta(r - q)/3 \tag{108}$$

from which

$$q = r - T_{zz}/2\eta; \; e = r - T_{zz}/3\eta; \; q/r = 1 - T_{zz}/3r\eta; \; \lambda^* = T_{zz}/r. \tag{109}$$

If inertia terms are neglected, there is no way of determining r and the problem is indeterminate. It becomes determinate if e vanishes, which gives $r = T_{zz}/3\eta$, $q/r = 0$ and $\lambda = 3\eta$. If the meaning of Trouton's experimental results is that in a viscous liquid in a first approximation $\lambda/\eta = 3$, this shows that Stokes' relation between stress and strain makes sense only if e is assumed to vanish. This does not mean that the liquid must be incompressible, but only that it is considered in a state when no change of volume takes place or after all such change has taken place.

7. We may now take up the general problem where we left it at the end of **4.** We shall, however, in the following deal with the second stage only, when e vanishes.

In viscous traction, while stresses may be high, the rate of strain is small. As a matter of fact this test was devised for such materials where large stresses produce small velocity strains only. In order to make use of this fact, we have to express λ in terms of *strain*.

For this we use (41), which gives, considering the second of equations (88), the second of equations (93) and (86) and (85),

$$T_{zz}\gamma'_s{}^r = F_1(I') \cdot 3/2 \cdot T_{zz}/\lambda \cdot \gamma'_s{}^r + F_2(I') \cdot 3/4 \cdot T^2{}_{zz}/\lambda^2 \cdot \gamma'_s{}^r \tag{110}$$

This yields

$$\lambda = 3/2 \cdot [F_1(0, I'_2, I'_3) + \tfrac{1}{2}F_2(0, I'_2, I'_3)e_{zz}]. \tag{111}$$

Considering (78), we now develop the functions F_1 and F_2 into power series of the argument I'_3 as follows

$$F_1(0, I'_2, I'_3) = 2(\eta + \eta_1 I'_3 + \eta_2 I'^2_3 + \cdots); \; F_2(0, I'_2, I'_3) = cI'_3 + c_1 I'^2_3 + \cdots \tag{112}$$

so that generally

$$T'^r_s = 2(\eta + \eta_1 I'_3 + \cdots)e'^r_s + (cI'_3 + \cdots)(e'^r_a e'^a_s + 2/3 \cdot I'_2 \cdot \delta^r_s) . \tag{113}$$

The coefficients η and c will generally be functions of the velocity strain components through the invariant I'_2. However, if we assume that these are all of the same order, or if they are constants, we may neglect the terms in I'^2_3 and higher powers and postulate a liquid which we may call the Trouton Liquid, with a rheological equation accordingly.

Its coefficient of viscous traction is from (111) and (112)

$$\lambda = 3\eta(1 + \eta_1/\eta \cdot I'_3 + c/4\eta \cdot I'_3 \cdot e_{zz}) \tag{114}$$

or introducing $I'_3 = e_{zz}^3/4$,

$$\lambda/\eta = 3(1 + 1/4 \cdot \eta_1/\eta \cdot e_{zz}^3 + 1/16 \cdot c/\eta \cdot e_{zz}^4) . \tag{115}$$

We have seen that for the simple Newtonian liquid, even should the coefficients of viscosity be not constant, λ/η cannot surpass the value 3. This is also the case with regard to the Reynolds liquid. Should, therefore, the average value of 3.16 found by Trouton be of significance, which it is difficult to judge, this would indicate the presence of the rheological coefficients η_1 or c or both. The term to which η_1 is attached changes its sign with the direction of the traction. In contradistinction the term to which c is attached does not do so. Experiments of viscous traction in *push* instead of pull, which Trouton did *not* carry out, would accordingly furnish criteria for the retention of one or the other of the two coefficients peculiar to the Trouton liquid.

8. Conclusions. The rheological equation of the most general Newtonian liquid so far suggested by experimental evidence is

$$T^{*r}_s = (3\kappa\epsilon + 3\eta_v e + \delta I'^r_s)\delta^r_s + 2(\eta + \eta_1 I'_3)e'^r_s + cI'_3(e'^r_a e'^a_s + 2/3 \cdot I'_2\delta^r_s)$$

which may be called the Trouton Liquid—where T^{*r}_s is the total stress (elastic and viscous), 3ϵ the cubical dilatation, $3e$ the rate thereof, e'^r_s the deviator of the velocity strain and κ, η_v, δ, η, η_1, c are six rheological coeffi-

cients. If a cylinder of the material is subjected to a simple axial traction T_{zz} and extends with the rate e_{zz}, the coefficient of viscous traction after a steady state has been reached, i. e., when $e = 0$, is

$$\lambda = T_{zz}/e_{zz} = 3\eta(1 + \eta_1/4\eta \cdot e_{zz}^3 + c/16\eta \cdot e_{zz}^4) \quad \text{(ii)}.$$

When η_1 and c vanish, the liquid is a Reynolds Liquid. The Trouton Liquid behaves in the same way as the Reynolds Liquid in tests for which I'_3 vanishes, e. g. in laminar flow. When, in addition, δ vanishes the liquid is a simple Newtonian Liquid. For cases of flow in which $e = 0$, the simple Newtonian Liquid is reduced to the Stokes Liquid. The rheological equation $T^{*}{}_s{}^r = -(p + 2\eta e)\delta_s{}^r + 2\eta e_s{}^r$ from which the Stokes-Navier differential equations in their general form are derived leads in viscous traction to unacceptable results. The Stokes-Navier differential equations should therefore be used in conjunction with $e = 0$ only.

THE STANDARDS INSTITUTION OF PALESTINE,
TEL AVIV.

For bibliography see first paper.

AN ARITHMETIC CHARACTERIZATION OF PROPER CHARACTERISTICS OF LINEAR SYSTEMS.*

By G. B. Huff.

Introduction. A *linear system* Σ of algebraic curves in a plane Π may be given by an equation of the form

$$\lambda_0 f_0 + \lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_d f_d = 0, \tag{1}$$

where $f_0, f_1, f_2, \cdots, f_d$ are $d+1$ linearly independent ternary polynomials of *order* x_0, and $\lambda_0, \lambda_1, \cdots, \lambda_d$ are parameters; d is said to be the *dimension* of the system. This article is concerned with linear systems which are defined by giving the order x_0 and the *multiplicities* $x_1, x_2, \cdots, x_\rho$ at a set of ρ general points $P_1, P_2, \cdots, P_\rho$ called the *base* of the system. If a system contains all the curves of order x_0 with the stipulated multiplicities at the base, it is said to be *complete.* If the conditions imposed by the base are independent, the system is said to be *regular.* Suppose, further, that at least one curve of the system is irreducible, has no multiple points outside the base, and is of genus p. For such a complete and regular linear system, its characteristic $x \equiv \{x_0; x_1, x_2, \cdots, x_\rho\}$ satisfies Cremona's equations [2]:

$$\begin{aligned} x_0^2 - x_1^2 - x_2^2 - \cdots - x_\rho^2 &= d + p - 1 \\ 3x_0 - x_1 - x_2 - \cdots - x_\rho &= d - p + 1. \end{aligned} \tag{2}$$

Since there is an irreducible curve in the system, a line through P_1, P_2 may not meet this curve in more than x_0 points; a conic through $P_1, P_2, \cdots, P_5$ may not meet the curve in more than $2x_0$ points, etc. Thus the characteristic of a linear system satisfies the inequalities [3]:

$$\begin{aligned} x_0 - x_1 - x_2 &\geq 0 \\ 2x_0 - x_1 - x_2 - \cdots x_5 &\geq 0 \\ 3x_0 - x_1 - x_2 - \cdots - x_9 &\geq 0 \\ 3x_0 - 2x_1 - x_2 - \cdots - x_7 &\geq 0 \\ \cdots\cdots\cdots\cdots & \end{aligned} \tag{3}$$

* Received April 3, 1946. Presented to the American Mathematical Society April 26, 1946.

It is understood that this system of inequalities includes all those of the form:

$$(3') \qquad y_0x_0 - y_1x_1 - y_2x_2 - \cdot \cdot \cdot - y_\rho x_\rho \geq 0 ,$$

where $y \equiv \{y_0; y_1, \cdot \cdot \cdot, y_\rho\}$ is the characteristic of a curve (linear system with $d = 0$) of lower order than x_0. If these inequalities hold for x so ordered that $x_1 \geq x_2 \geq \cdot \cdot \cdot \geq x_\rho$ and with all the characteristics y similarly ordered, it is clear that they hold *a fortiori* for all other orderings.

In 1934, Coble [2] gave a determination of all the integer solutions of Cremona's equations (2) for any given values of p, d, ρ. He used the word characteristic for any solution of (2) and characteristics were classified as *proper, degenerate* or *virtual* according as the general curve of the system is existent and irreducible, existent but reducible, or non-existent.

These results underlined the need for arithmetic criteria for proper characteristics. In this paper that problem is investigated and it is in fact shown that, *for early values of* p, d, *every set of integers satisfying* (2) *and* (3′) *is the characteristic of a linear system.*

A basic tool is the notion of the image of a characteristic x under A_{123}. Let a quadratic transformation be set up between the plane Π and a plane Π', with fundamental points at P_1, P_2, P_3 in Π and call the fundamental points in Π', P_1', P_2', P_3'. Designate the images of $P_4, P_5, \cdot \cdot \cdot, P_\rho$ in Π' by P_4', $P_5', \cdot \cdot \cdot, P_\rho'$. The image of Σ is a linear system Σ' in Π', determined by a characteristic x' at the base $P_1', P_2', P_3', \cdot \cdot \cdot, P_\rho'$; and x' is given by

$$(4) \qquad A_{123}: \quad \begin{array}{ll} x_i' = x_i + (x_0 - x_1 - x_2 - x_3) & i = 0, 1, 2, 3 \\ x_j' = x_j & j = 4, 5, \cdot \cdot \cdot, \rho. \end{array}$$

Given a proper characteristic, the linear transformation A_{123} may be used to construct a new proper characteristic; moreover, if x' is proper, it follows that x is proper.

Since the points of a base may be numbered at will, the image of a characteristic x under any permutation of $x_1, x_2, \cdot \cdot \cdot, x_\rho$ is proper if and only if x is. Thus A_{123} and the interchanges (x_1x_2), $(x_1x_3), \cdot \cdot \cdot, (x_1x_\rho)$ generate a linear transformation group, $\mathfrak{G}_\rho$, with the property that if x is a proper characteristic and $S \,\varepsilon\, \mathfrak{G}_\rho$, then $x' = S(x)$ is proper. The bilinear form $y_0x_0 - y_1x_1 - \cdot \cdot \cdot - y_\rho x_\rho$ of the inequalities (3′) is invariant under A_{123} and hence for any $S \,\varepsilon\, \mathfrak{G}_\rho$.

The line L through P_1, P_2, a linear system of $p = d = 0$ and charac

teristic $l = \{1; 1, 1, 0, \cdots, 0\}$ is a principal curve for the quadratic transformation with fundamental points at P_1, P_2, P_3. Its image under this transformation is the set of directions about P_3'. The image of l under A_{123} is $\{0; 0, 0, -1, 0, \cdots, 0\}$. This special characteristic is defined to be proper and to represent the directions about P_3'.

1. The inequality $x_0 - x_1 - x_2 - x_3 \geqq 0$. Two characteristics x, x' are said to be of the same *type* if x' is the image of x under a permutation of $x_1, x_2, \cdots, x_\rho$. If x has the property $x_0' - x_1' - x_2' - x_3' \geqq 0$ for all x' of the same type, it is clear that its image under A_{123} is of the same or higher order, and that the same would be true for any x' of the same type. Such a characteristic is said to be of *minimal order*. Noether [7], Bertini [1], Guccia [4], and others have obtained a variety of results along these lines. Since in each of these papers the characteristics were all assumed to be proper, the theorems as stated are not directly applicable here. The arithmetical results were obtained on the assumption that $x_0, x_1, \cdots, x_\rho$ were non-negative. If we call such a characteristic non-negative, the results may be stated as follows:

(5) (Noether) *The only non-negative characteristic of minimal order for $p = 0$, $d = 2$, is $\{1; 0, 0, \cdots, 0\}$.*

(6) (Noether) *The only non-negative characteristics of minimal order for $p = 0$, $d = 1$ are of type $\{1; 1, 0, \cdots, 0\}$.*

(7) (Bertini) *There is no non-negative characteristic of minimal order for $p = d = 0$, but those of type $\{0; 0, \cdots, 0, -1\}$ are of minimal order.*

(8) (Bertini-Guccia) *The only non-negative characteristics of minimal order for $p = 1$, $d = 0$ are of type $\{3\mu; \mu^9\, 0^{\rho-9}\}$.*

(9) (Guccia) *The only non-negative characteristics of minimal order for $p = 1$, $d = 1, 2, \cdots, 7$ are of type $\{3; 1^{9-d}, 0, \cdots, 0\}$.*

These are readily proved without reference to the original papers by the following device, (which is essentially what all these early workers used): in (2) set $x_0 = x_1 + x_2 + x_3 + \delta$ and then subtract x_3 times the second from the first, getting:

$$(*)\quad \delta^2 + (2x_1 + 2x_2 - x_3)\delta + 2(x_1x_2 - x_3^2) + x_4(x_3 - x_4) + \cdots + x_\rho(x_3 - x_\rho) = (d + p - 1) - x_3(d - p + 1).$$

Now if x is non-negative and ordered so that $x_1 \geq x_2 \geq \cdots \geq x^\rho$, the results $(5), \cdots, (9)$ are easily established by substituting the values of p, d in $(*)$ and making quite elementary arguments.

From the arithmetic proposition (7) Bertini obtained a geometric theorem which will be stated as follows for use in the next section:

(10) *If l is a proper solution of (2) for $p = d = 0$, there exists an $S \varepsilon \mathfrak{G}_\rho$ (i. e., an S which is the product of transformations A_{123} and permutations) such that $S(l) = \{0; 0, 0, \cdots, 0, -1\}$.*

The curve defined by a proper l is a rational curve determined by its behavior at a base, a Bertini L-curve. The totality of L-curves for a given base is the same as the collection of all principal curves of homaloidal nets defined at that base.

2. Property A and arithmetic inequalities. In a recent article [6] this writer studied characteristics x with $x_0 > 0$ and p, d non-negative. This was designated as property A. Of the results obtained there, the following ones are needed.

(11) *If x^* is obtained from x by setting $x_\rho = 0$, then*

$$p^* = p + x_\rho(x_\rho - 1)/2, \qquad d^* = p + x_\rho(x_\rho + 1)/2.$$

If x has property A, then so does x^*.

(12) *If x has property A and c is the characteristic of a homaloidal net, then $c_0x_0 - c_1x_1 - c_2x_2 - \cdots - c_\rho x_\rho \geq 0$. Moreover, the equality sign holds only for the characteristics of the principal curves of the homaloidal net.*

These will be used in establishing the theorem below, which is a substantial generalization of Theorem 3 of the earlier paper [6].

(13) *If x has property A and l is a proper characteristic of $p = d = 0$ and $l_0x_0 - l_1x_1 - l_2x_2 - \cdots - l_\rho x_\rho < 0$, then $x_0 \geq l_0$. Moreover, $x_0 = l_0$ only if $x = l$.*

If x is a proper characteristic of $p = d = 0$, the case is quickly settled.

x and l both represent irreducible Bertini L-curves which meet at more than $x_0 l_0$ points and coincide. Thus $x = l$.

If x is not the characteristic of a Bertini L-curve, the reasoning goes as follows. Since l is proper, according to (10) there exists an $S \varepsilon \mathfrak{G}_\rho$ such that $\bar{l} = S(l) = \{0; 0, \cdots, 0, -1\}$. From the fact that x is not the characteristic of a Bertini L-curve, it is clear from (12) that $\bar{x} = S(x)$ has $\bar{x}_0 > 0$. The invariance of the bilinear form $y_0 x_0 - y_1 x_1 - y_2 x_2 - \cdots - y_\rho x_\rho$ under S implies that

$$l_0 x_0 - l_1 x_1 - \cdots - l_\rho x_\rho = \bar{l}_0 \bar{x}_0 - \bar{l}_1 \bar{x}_1 - \cdots - \bar{l}_\rho \bar{x}_\rho = \bar{x}_\rho,$$

and $\bar{x}_\rho$ is negative, say $\bar{x}_\rho = -k$, $k > 0$. Thus $\bar{x}$ may be written in the form

$$\bar{x} = \bar{x}^* + k\bar{l},$$

where $\bar{x}^*$ is $\bar{x}$ with $\bar{x}_\rho$ replaced by zero and k is a positive integer. Consider now the image of $\bar{x}$ under S^{-1}.

$$S^{-1}(\bar{x}) = S^{-1}(\bar{x}^* + k\bar{l}) = S^{-1}(\bar{x}^*) + kS^{-1}(\bar{l}),$$

or

$$x = (\bar{x}^*)' + kl.$$

By (11), $\bar{x}^*$ has $p^* + d^* = p + d + k^2 > 0$ and by (12), $(\bar{x}_0{}^*)' > 0$. From

$$x_0 = (\bar{x}_0{}^*)' + kl_0,$$

it follows that

$$x_0 > l_0.$$

This result is put into the following form to make it more usable in the next proof.

(14) *If l is a proper characteristic of $p = d = 0$, and x is any other characteristic of property A and $x_0 \leq l_0$, then $l_0 x_0 - l_1 x_1 - l_2 x_2 - \cdots - l_\rho x_\rho \geq 0$.*

Any characteristic that satisfies the inequalities (3′) and has $x_1, x_2, \cdots, x_\rho$ non-negative satisfies $l_0 x_0 - l_1 x_1 - l_2 x_2 - \cdots - l_\rho x_\rho \geq 0$ for all proper characteristics l of $p = d = 0$ and $l_0 < x_0$, including in particular the special

proper characteristics of type $\{0;0,0,\cdots,0,-1\}$. For such an x the following statement may be made.*

THEOREM I. *If x has property A and if*

$$l_0x_0-l_1x_1-l_2x_2-\cdots-l_\rho x_\rho\geqq 0$$

for all proper l of $p=d=0$ and $l_0<x_0$; and x' is the image of x under any $S\,\varepsilon\,\mathfrak{G}_\rho$, then either $x_0'>0$ and $x_i\geqq 0$; or x' is of type $\{0;0,0,\cdots,0,-1\}$

Suppose that x is not a proper characteristic of $p=d=0$. The hypothesis of the theorem and (14) assure us that $l_0x_0-l_1x_1-l_2x_2-\cdots-l_\rho x_\rho\geqq 0$ for all proper l. Now the set $\{l\}$ is merely permuted by any $S\,\varepsilon\,\mathfrak{G}_\rho$ and the bilinear relation $l_0x_0-l_1x_1-\cdots-l_\rho x_\rho$ is invariant under S. Thus $l_0x_0'-l_1x_1'-\cdots-l_\rho x_\rho'\geqq 0$ for all proper l. Since these include those of type $\{0;0,0,\cdots,0,-1\}$, it follows that $x_i\geqq 0$. By (12), $x_0'>0$.

If x is indeed a proper characteristic of $p=d=0$, its image under any $S\,\varepsilon\,\mathfrak{G}_\rho$ is proper and must have $x_0'>0$, $x_i'\geqq 0$ or must be of type $\{0;0,0,\cdots,0,-1\}$.

3. Arithmetic criteria for proper characteristics.

In this section it will be shown that simple arithmetic conditions exist which assure the properness of characteristics for eleven values of p, d. Two of the cases are somewhat different from the others and are treated separately.

THEOREM II. *If x is a non-negative characteristic of $p=d=0$; such that $l_0x_0-l_1x_1-\cdots-l_\rho x_\rho\geqq 0$ for all proper characteristics l of $p=d=0$ and $l_0<x_0$, then x is proper.*

By Theorem (7), x is not of minimal order. Let p be the permutation that sends the three highest multiplicities into x_1, x_2, x_3 and let $x'=A_{123}p(x)$. It is clear that $x_0>x_0'$. By Theorem I, either x_0' is of type $\{0;0,0,\cdots,0,-1\}$ or it is non-negative. In the first case x' is proper and in the second x' is not of minimal order and the process may be continued. Since $x_0>x_0'>x_0''>\cdots$ is a sequence of non-negative integers, there must finally result some image x^* of type $\{0;0,0,\cdots,0,-1\}$. Thus x is proper.

An entirely similar argument yields:

* The author is indebted to R. J. Walker for the conjecture that this theorem might be valid.

THEOREM III. *If x is a non-negative characteristic of p, $d=1,0$ such that the g. c. d. of $x_0, x_1, \cdots, x_\rho$ is 1 and $l_0x_0 - l_1x_1 - l_2x_2 - \cdots - l_\rho x_\rho \geqq 0$ for all proper l of $p=d=0$ and $l_0 < x_0$, then x is proper.*

Now by Theorem (8), x is of type $\{3\mu; \mu^9, 0, \cdots, 0\}$ or is not of minimal order. In the first case $\mu = 1$ and x is proper. As before let p be a permutation such that $p(x)$ does not satisfy $x_0 - x_1 - x_2 - x_3 \geqq 0$ and consider $x' = A_{123}p(x)$. Again $x_0 > x_0'$ and either $x' = \{3\mu; \mu^9, 0, \cdots, 0\}$ or it is not of minimal order. Thus a sequence of characteristics $x, x', \cdots, x^*$ can be set up so that $x_0 > x_0' > \cdots > x_0^*$, where $x^* = \{3\mu; \mu^9, 0, \cdots, 0\}$. But the g. c. d. of $x_0^*, x_1^*, \cdots, x_\rho^*$ is 1, and hence $x^* = \{3; 1^9, 0, \cdots, 0\}$ and x is proper.

These arguments indicate clearly the method of proof to be used in establishing:

THEOREM IV. *If x is a characteristic of $p, d = 0, 1; 0, 2; 1, d$ $(d = 1, \cdots, 7)$ such that $l_0x_0 - l_1x_1 - \cdots - l_\rho x_\rho \geqq 0$ for all proper l of $p = d = 0$ and $l_0 < x_0$, then x is proper.*

These theorems are all actually necessary and sufficient condition theorems, since proper positive ordered characteristics of these sorts must satisfy the inequalities (3). However, the interesting thing is the sufficiency part of the theorems.

Conclusion. The question raised in the introduction must then be answered in the affirmative for the values of p, d considered in Section **3**. Any characteristic x which satisfies (2) and (3′) for those values of p, d must be proper. Indeed, it is sufficient that x satisfy (3′) for the characteristics of Bertini L-curves. This result was already known [5] for $p=0$, $d=2$, but was, in a sense, vacuous since there existed no arithmetic criterion for $p=d=0$. Thus the validity of Theorem II adds to the significance of the other results.

To construct arithmetically a table of proper characteristics of Bertini L-curves $(p=d=0)$, classified according to order x_0, one might proceed as follows. Suppose that the table has been computed up to $x_0 = n-1$. Then for the order n, there is a finite number of solutions of (2) in non-negative integers. All these could be tested in terms of the earlier ones. The process could be carried as far as desired with (a) certainty of getting all proper

characteristics of each order and (b) no danger of including improper cha teristics. This table could then be used to check construction of tabl other values of p, d.

REFERENCES.

(1) E. Bertini, "Sulle curve razionali," *Giornale di Mat.*, ser. (1), vol. 15, pp. 32

(2) A. B. Coble, "Cremona's diophantine equations," *American Journal of M matics*, vol. 56 (1934), pp. 459-489.

(3) ———, *American Mathematical Society Colloquium Publications*, vol. 10, p.

(4) G. B. Guccia, "Generalizzazione di un teorema di Nöther," *Rendiconti del C Mathematico di Palermo*, vol. 1 (1884), pp. 139-156.

(5) G. B. Huff, "A sufficient condition that a C-characteristic be geometric *ceedings of the National Academy of Sciences*, vol. 29 (1943), pp. 19

(6) ———, "Inequalities connecting solutions of Cremona's equations," *Bull the American Mathematical Society*, vol. 52 (1946), pp. 287-291.

(7) M. Noether, "Ueber Flächen, welche Schaaren rationaler Curven be *Mathematische Annalen*, vol. 3 (1871), pp. 161-227.